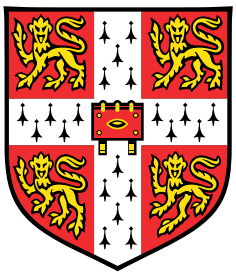
Robust methods in Mendelian randomization



Jessica Mary Barbara Rees

Department of Public Health and Primary Care
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

St Catherine's College

January 2019

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Acknowledgements and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Acknowledgements and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Acknowledgements and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Jessica Mary Barbara Rees
January 2019

Robust methods in Mendelian randomization

Jessica Mary Barbara Rees

Mendelian randomization uses genetic variants as instrumental variables to estimate the causal effect of a risk factor on an outcome using observational data. If a genetic variant is included in a Mendelian randomization study that does not satisfy the instrumental variable assumptions then the causal estimate from traditional instrumental variable methods will be biased. Since Mendelian randomization studies using publicly available summary level data (estimates and standard errors of the genetic associations with the risk factor and the outcome) from large consortia can be performed with relative ease and little expense, the popularity of Mendelian randomization in epidemiological studies has increased dramatically. As such, various methods have been developed in Mendelian randomization that use summary level data and account for possible violations of the instrumental variable assumptions. However, additional Mendelian randomization methods that account for violations in the instrumental variable assumptions are still required.

In this dissertation, we introduce robust methods for Mendelian randomization that downweight the contribution of genetic variants with heterogeneous causal ratio estimates. We extend the univariable MR-Egger method to the multivariable setting to account for both measured and unmeasured pleiotropic effects. We also explore the possibility of extending multivariable Mendelian randomization to the factorial setting to estimate statistical interaction effects. Finally, we apply some of the methods we have developed to perform a Mendelian randomization study to investigate the effect of adiposity and body composition measurements on asthma using data from UK Biobank and the GABRIEL consortium.

This dissertation is dedicated to my uncle, Canon Peter Coyle, who passed away during my PhD. He will always be remembered for his kindness, humility, and fortitude.

Acknowledgements

I would like to thank my supervisor, Stephen Burgess, for his guidance and support, and my secondary supervisor, Angela Wood, for advice and feedback on the work presented in this dissertation. I would also like to thank Chris Foley for helpful discussions and contributions to the work presented in Chapter 5, and Raquel Granell for feedback on the work in Chapter 6.

The work presented in Chapter 3 is based on work originally carried out by my supervisor Stephen Burgess and other collaborators (Jack Bowden, Frank Dudbridge and Simon Thompson). A copy of a draft manuscript that has been uploaded to arXiv on this work by Stephen Burgess and colleagues can be found in Appendix A. I acknowledge that Stephen Burgess developed the methods in Sections 3.3.1 to Section 3.3.3. I suggested the extension in Section 3.3.4, performed the additional applied analysis in Section 3.4, adapted and re-performed the simulation studies, and re-wrote the work. I would like to thank Stephen Burgess, Angela Wood, Jack Bowden and Frank Dudbridge for their useful comments on the work contained in Chapter 3. The paper in Appendix B is based on the work in Chapter 3 and was written by me under the supervision of Stephen Burgess, with contributions from Angela Wood, Jack Bowden and Frank Dudbridge.

The paper in Appendix D was written by me under the supervision of Stephen Burgess, with contributions and editorial input from Angela Wood. This paper forms the basis of the work in Chapter 4.

The paper in Appendix F was written by me under the supervision of Stephen Burgess, with contributions from Chris Foley. This paper forms the basis of the work in Chapter 5.

Finally, I would like to thank my family and friends for all their help and encouragement. I would particularly like to acknowledge the unwavering support shown by my parents, Andrew and Margaret.

Abbreviations

AD	Alzheimer's disease
ATE	average treatment effect
BIA	bioelectrical impedance analysis
BMI	body mass index
CHD	coronary heart disease
CI	confidence interval
DAG	directed acyclic graph
DXA	dual X-ray emission absorptiometry
FEV1	forced expiratory volume in 1 second
FFM/FFMI	fat free mass/fat free mass index
FM/FMI	fat mass/fat mass index
FVC	forced volume vital capacity
GABRIEL	Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community
GIANT	Genetic Investigation of ANthropometric Traits
GLIDE	global and individual tests for direct effects
GLGC	Global Lipids Genetics Consortium
GWAS	genome-wide association study
GWS	genome wide significance
HDL-C	high-density lipoprotein cholesterol
InSIDE	instrument strength independent of direct effect
ITT	intention to treat
IV	instrumental variable
IVW	inverse-variance weighted
LDL-C	low-density lipoprotein cholesterol
LM/LMI	lean mass/lean mass index
LP	Lasso penalization
LTS	least trimmed squares
MAF	minor allele frequency
MBE	mode-based estimator
MR-Egger	Mendelian randomization-Egger
NOME	no measurement error
OLS	ordinary least squares
OR	odds ratio
PC	principle component
PW	penalized weights
QC	quality control
RCT	randomized clinical trial
Rr	robust regression
RR	relative risk
SBP	systolic blood pressure

SD	standard deviation
SE	standard error
SNP	single nucleotide polymorphism
SSGAC	Social Science Genetic Association Consortium
TAG	Tobacco, Alcohol and Genetics
TSLS	two-stage least squares
ZEMPA	zero modal pleiotropy assumption

Notation

X	risk factor
$\bar{X}$	mean value of the risk factor (Chapter 5)
Y	outcome
U	confounder of the $X - Y$ association
Z	instrumental variable
G	genetic variant acting as an instrumental variable
UI	univariable inverse-variance weighted (Chapter 4)
UE	univariable MR-Egger (Chapter 4)
MI	multivariable inverse-variance weighted (Chapter 4)
ME	multivariable MR-Egger (Chapter 4)
GS	weighted gene score
α	direct effect of G on Y
α'	direct effect of G on Y in a multivariable framework (Chapter 4)
β_X	parameter of genetic association: regression parameter in the $G - X$ regression
β_Y	parameter of genetic association: regression parameter in the $G - Y$ regression
γ	causal effect of one risk factor on another risk factor (Chapter 4)
δ	parameter of genetic association: regression parameter in the $G - U$ regression (Chapter 3)
ζ_X	parameter of genetic association: regression parameter in the $U - X$ regression
ζ_Y	parameter of genetic association: regression parameter in the $U - Y$ regression
ϵ	error term in linear regression models
θ	causal effect of X on Y
θ_M	marginal causal effect of X on Y (Chapter 5)
λ	penalty term in Lasso regression (Chapter 3)
μ	mean value of α' (Chapter 4) or mean value of X (Chapter 5)
ρ	correlation parameter
σ	variance parameter
ϕ	random effects parameter
F	F statistic from regression of X on G
h	number of data points in least trimmed squares (Chapter 3)
i	subscript indexing individual
j	subscript indexing genetic variant
J	total number of genetic variants
k	subscript indexing risk factor (Chapter 4)
K	total number of risk factors (Chapter 4)
N	total number of individuals or observations
n	number of increments in the heterogeneity stopping rule (Chapter 3)

p	probability of Y (Chapter 5)
Q	Cochran's Q statistic
r	standardised residuals (Chapter 3)
r^2	measure of linkage disequilibrium (Chapter 6)
R^2	amount of variance explained in X
w	inverse-variance of the causal ratio estimate
w_{LTS}	weights from least trimmed squares regression (Chapter 3)
$\mathcal{B}$	binomial distribution
$\mathcal{N}$	normal distribution
$\mathcal{U}$	uniform distribution
χ^2	chi-squared distribution

Note that bold variables represent vectors.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Epidemiology investigates the determinants of health outcomes and the distribution of diseases at the population level. Epidemiological research can inform disease aetiology, the effectiveness of treatments on disease outcomes, public health policies, and the prioritization of healthcare resources. To strengthen the credibility of these findings and recommendations, epidemiology must consider questions of cause and effect.

In this Chapter, we discuss what is meant by the causal effect of a risk factor on an outcome (Section 1.1), and consider the detection and estimation of these effects in randomized clinical trials (Section 1.2), epidemiological studies with observational data (Section 1.3), instrumental variable analyses (Section 1.4), and finally, Mendelian randomization (Section 1.5). We then provide motivation for the work presented (Section 1.6), and outline the structure of the dissertation (Section 1.7).

1.1 Causal effect

In this dissertation, we consider a causal effect to be a measure of the impact an intervention on the risk factor X has on the distribution of an outcome Y. We use the notation $do(X = x)$ introduced by Pearl [1] to illustrate that X has been intervened on and set to the value x. If the risk factor has a causal effect on the outcome, then an intervention on X will change the distribution of Y, and the conditional distribution $p(Y = y|do(X = x))$ will be dependent on the value x.

An observational association considers differences in the outcome when the risk factor is observed at different values. If the risk factor is associated with other variables, such as variables that confound the association between X and Y, then the observational association of the risk factor on the outcome may highlight differences in the risk factor and the confounders. Hence, the conditional distribution $p(Y = y|X = x)$ may not

be equivalent to $p(Y = y|do(X = x))$ [2]. This distinction between an observational association and a causal effect has contributed to the phrase 'correlation does not imply causation'.

1.2 Randomized clinical trials

Randomized clinical trials (RCT) are considered the 'gold standard' of assessing the effectiveness of a treatment on a disease outcome [3]. In its simplest form, a RCT randomly allocates participants to receive the treatment or control (no treatment). By randomizing participants to treatment, all known and unknown confounders should be balanced between the two treatment groups [4]. It can be inferred that the treatment has a causal effect on the outcome if the frequency of the disease outcome differs between the randomized groups. RCTs usually perform an intention to treat (ITT) analysis where the effect of randomization on the outcome is estimated [5]. The estimate from a ITT analysis will be equivalent to the causal parameter of the average treatment effect (ATE) if all of the participants take the treatment they have been randomly allocated to ('full compliance to randomization') [6].

1.3 Epidemiological studies

Due to cost, time, and ethical reasons, it may not be feasible to conduct a RCT to investigate the causal effect of a treatment or modifiable risk factor on a disease outcome [7]. Observational data is often used to investigate the effect of a risk factor on a disease outcome when a RCT cannot be performed. Whilst it is theoretically possible to adjust for the confounders of the risk factor–outcome association in the statistical analysis of the observational data, we cannot guarantee that all of the confounders will have been accounted for (known as 'residual confounding') [8]. Observational studies may also be affected by 'reverse causation' in which the observed association between the risk factor and outcome is due to a causal effect of the outcome on the risk factor [8].

Due to residual confounding and reverse causation, analyses using observational data that adjust for potential confounders in the statistical model cannot distinguish between correlation and causation. This limitation has led to numerous examples where an apparent association has been identified using observational data, but the result has not been replicated in a RCT. For example, epidemiological studies using observational data suggested that vitamin C has a protective effect against cardiovascular disease

[9, 10], but this result was not supported in a RCT where a null effect was reported [11].

1.4 Instrumental variable analyses

Using observational data, an instrumental variable (IV) can be used to infer a causal effect between a risk factor X and an outcome Y. IVs have been applied to a wide range of research areas, including economics and medical research. For a variable G to be a valid IV, the following conditions must be satisfied:

- IV1: G is associated with the risk factor ($G \not\perp\!\!\!\perp X$);
- IV2: G is (marginally) independent of all unmeasured confounders U of the risk factor–outcome association ($G \perp\!\!\!\perp U$); and
- IV3: G is independent of the outcome conditional on the risk factor and confounders ($G \perp\!\!\!\perp Y|(X, U)$) [2, 12].

Under the IV1 assumption, there will be a systematic difference in the average levels of the risk factor between the subgroups of G, and the IV2 assumption ensures that the unmeasured confounders U will be equally distributed between these subgroups. The IV3 assumption guarantees that G only has an effect on Y via X, i.e. G does not have a direct effect on Y.

Figure 1.1 is a directed acyclic graph (DAG) of the variables G, X, Y and U, where G satisfies the IV assumptions. A DAG is a graphical model that provides a non-parametric representation of the relationships between a set of variables. In a DAG, nodes are used to represent variables, and these nodes are connected by directed edges, usually represented as single headed arrows. For example, $A \rightarrow B$ implies that the variable A has a direct effect on variable B. Conversely, if two nodes are not connected by an arrow, this infers that there is no direct effect between the two variables. A DAG must not contain a variable C which has a sequence of directed edges that lead back to C. Hence, a DAG cannot have any complete cycles. DAGs do not have to contain all intermediate variables, such as C in $A \rightarrow C \rightarrow B$, and a node may represent a collection of variables.

In Figure 1.1, the IV1 assumption is satisfied by the arrow connecting G to the risk factor X. IV2 is satisfied as there is no arrow that directly links G and the set of unmeasured confounders U, and there is no pathway between G and U. Finally, IV3 is satisfied as the only pathway between G and the outcome Y is via the risk factor X.

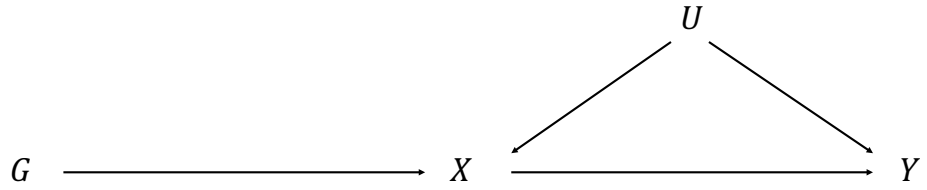


Fig. 1.1 Directed acyclic graph illustrating the instrumental variable assumptions for the variable G to investigate the causal effect of a risk factor X on a outcome Y, where U are the set of unmeasured confounding variables of the $X - Y$ association.

DAGs imply that there is a direction of effect between variables, but these effects are not necessarily causal. From Figure 1.1, the joint distribution of Y, X, U and G can be factorized as:

$$p(y, x, u, g) = p(y|u, x)p(x|u, g)p(u)p(g)\,, \tag{1.1}$$

and the DAG will have a causal interpretation with respect to X if we can intervene on X without changing the distributions $p(y|u, x)$, $p(u)$ and $p(g)$ in Equation 1.1 [13]. These three distributions should be the same regardless of whether X is set to x' (i.e. $do(X = x')$) or x' is observed. Sheehan and Didelez [13] refer to this condition as the 'structural assumption', and this assumption, along with the three IV conditions, must be satisfied for the DAG in Figure 1.1 to have a causal interpretation with respect to X. Under this structural assumption, we can express the joint distribution of Y, X, U and G as:

$$p(y, u, g|do(X = x')) = p(y|u, x')p(u)p(g)\,. \tag{1.2}$$

In Section 2.1, we consider how an IV can be used to statistically test for a causal effect between X and Y within the context of Mendelian randomization. We define the additional assumptions required to produce a point estimate of the causal effect (Section 2.2), and outline the methods commonly used in the IV literature to estimate the causal effect (Section 2.3).

1.5 Mendelian randomization

Mendelian randomization uses genetic variants as IVs to detect and/or estimate the causal effect of a risk factor on an outcome using observational data. Katan [14] first introduced the idea of using genetic variants as IVs to detect causal effects, and their use in epidemiological research has been popularized by Davey Smith and Ebrahim [15].

In this Section, we discuss the merits of using genetic variants as IVs and introduce different types of Mendelian randomization studies.

1.5.1 Using genetic variants as instrumental variables

A genetic variant must be associated with the risk factor for the IV1 assumption to be satisfied. Since there has been a substantial increase in the number of genome wide association studies (GWAS), and the results from these studies are usually publicly available, this assumption should be relatively straight forward to verify. Typically, uncorrelated genetic variants (not in linkage disequilibrium) that are associated with the risk factor at the genome wide significance level (p-value $< 5 \times 10^{-8}$) are considered in a Mendelian randomization study.

Since increases in sample sizes have led to more genetic variants being identified in GWASs, and common genetic variants typically explain little variation in the risk factor, many Mendelian randomization analyses now include multiple genetic variants as IVs [16]. The genetic variants do not have to be causally associated with the risk factor to be valid IVs. Any genetic variant that is in linkage disequilibrium with the causal variant and satisfies the IV assumptions can be used as a IV [17]. Including multiple genetic variants in the analysis will only increase the power to detect the causal effect if the variants explain additional variability in the risk factor [18, 19]. Note that since genetic variants are determined at conception, the association between the variant and the risk factor should not be subject to reverse causation [20, 21].

The IV2 assumption that the genetic variant is not associated with any of the unmeasured confounders of the risk factor–outcome association is an untestable condition. The assumption that genetic variants are 'randomly' distributed in the population, combined with Mendel's laws of inheritance, are often used to justify the validity of the IV2 assumption as it implies that the genetic variants are randomly distributed in the population with respect to potentially confounding variables, such as social and environmental factors [15]. The credibility of the IV2 condition could be considered by testing the genetic variants with known measured confounders of the risk factor–outcome association in the dataset used in the main analysis, and by looking up the genetic associations with known unmeasured confounders in external datasets and consortia. Although this is a sensible suggestion, it is by no means exhaustive.

If a genetic variant is associated with more than one trait then it is said to be a 'pleiotropic' variant. The inclusion of a pleiotropic genetic variant in a Mendelian randomization analysis may lead to the violation of the IV2 or IV3 assumptions. Since GWASs have identified many genetic variants that are associated with multiple traits,

including pleiotropic variants in a Mendelian randomization study is a major concern [22]. This limitation has led to various methods being introduced into the Mendelian randomization literature that either detect and remove pleiotropic variants, or estimate consistent causal effects in the presence of pleiotropic variants.

1.5.2 Classification of studies

Figure 1.2 provides an illustration of the two main types of Mendelian randomization studies considered in the literature and this dissertation, and the type of data that can be used in the analysis of these two studies. When Mendelian randomization was initially considered in the literature, data on the same set of individuals was generally used, known as a 'one–sample' Mendelian randomization study [23]. Typically, individual level data on the risk factor, outcome, and genetic variants are used in the analysis model for one–sample Mendelian randomization. However, it is possible for estimates and standard errors of the genetic associations with the risk factor and with the outcome, referred to as 'summary level data', to be used in the analysis of a one–sample Mendelian randomization study.

It has now become increasingly popular for Mendelian randomization analyses to use data from two independent samples, known as a 'two–sample' Mendelian randomization study [24]. Two–sample Mendelian randomization studies generally use summary level data where the estimates and standard errors of the genetic associations with the risk factor are obtained from one sample, and the estimates and standard errors of the genetic associations with the outcome are obtained from the other sample. It is assumed that the two independent samples come from the same underlying population.

Typically, 'summary level data' refers to the case where the genetic associations with the risk factor and the genetic associations with the outcome have been estimated in two independent samples (i.e. a two–sample Mendelian randomization study). However, as noted above, it is possible for summary level data to be used in a one–sample study. Throughout this dissertation, we assume that 'summary level data' refers to the two–sample setting unless explicitly stated otherwise.

Since access to individual level data can be restrictive, and summary level data is often publicly available from GWASs and large consortia, two–sample Mendelian randomization studies continue to grow in popularity [25]. This has led to numerous methodological developments in using summary level data in Mendelian randomization. Databases, such as Phenoscanner [26], and software, such as MR-Base [27], have been developed to allow users to extract summary level data from published GWASs and

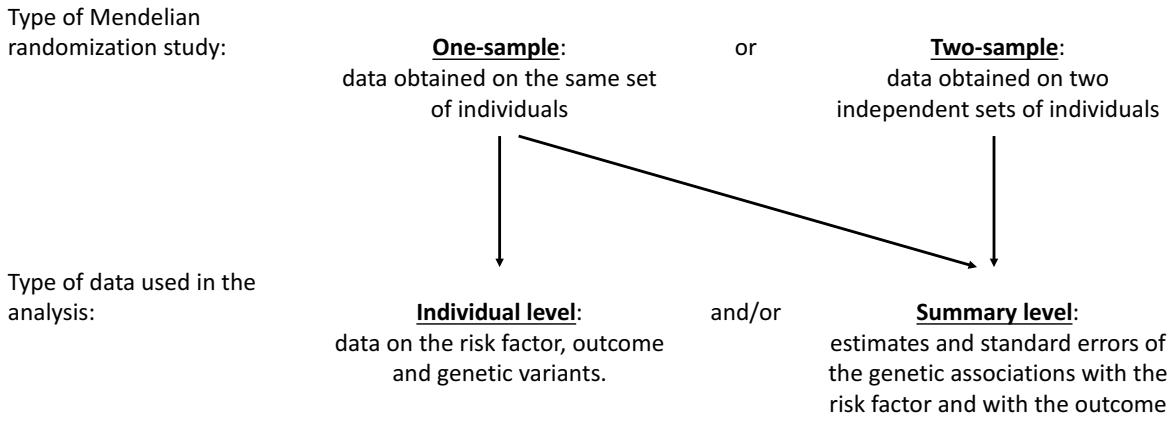


Fig. 1.2 Diagram illustrating some of the types of Mendelian randomization studies and the data used in the analysis.

consortia databases. MR-Base will even perform a two-sample Mendelian randomization analysis if the user specifies a risk factor and outcome.

1.6 Motivation for the dissertation

The overarching aim for this dissertation was to develop methods for applied Mendelian randomization studies. Some of the method development in this dissertation has been motivated by our main applied example of investigating the effect of adiposity and body composition on asthma using data from UK Biobank in a one–sample Mendelian randomization study. UK Biobank is a prospective, population based cohort consisting of approximately 500,000 participants aged between 40-69 years living in the UK. Extensive baseline characteristics were collected at recruitment, including adiposity, body composition measurements and genetic information.

The primary research question for this applied example was to use body mass index (BMI) as a measure of adiposity to perform a Mendelian randomization study to investigate the effect of BMI on asthma. Since the results from a Mendelian randomization study may be invalid if the IV2 or IV3 assumptions are violated, we consulted the literature on Mendelian randomization to identify methods that detect or account for the inclusion of pleiotropic variants (discussed in Chapter 2). Identifying gaps in the literature, we developed and extended 'robust methods' that downweight or remove pleiotropic genetic variants (Chapter 3). It was anticipated that these methods would be used in the sensitivity analysis for the Mendelian randomization study on BMI and asthma (Chapter 6).

UK Biobank has measurements on body composition, including fat mass (FM) and fat-free mass (FFM). To have a more comprehensive appreciation for the effect of adiposity and body composition on asthma, we considered the possibility of investigating the simultaneous effect of FM and FFM on asthma in a one–sample Mendelian randomization study. Since there is a substantial overlap in the genetic variants that are associated with FM and FFM, the IV assumptions would have been violated if the effect of FM and FFM on asthma had been considered in separate Mendelian randomization analyses. However, multivariable Mendelian randomization has been developed to allow for causal effects of multiple risk factors that share common genetic predictors to be estimated in the same analysis [28].

The estimates from multivariable Mendelian randomization may be invalid if pleiotropic variants that are associated with traits that do not lie on the causal pathways between the risk factors and outcome (violation of the IV3 assumption for multivariable Mendelian randomization) are included in the analysis. However, there are no methods that estimate consistent causal effects of multiple risk factors when pleiotropic variants that violate the IV3 assumption for multivariable Mendelian randomization are included in the analysis. Since the MR-Egger method has been developed to estimate consistent causal effects in the presence of pleiotropic variants when one risk factor is included in the analysis [29], we considered the extension of this method to the multivariable setting (Chapter 4) with the anticipation of using the method in the sensitivity analysis for investigating the effect of FM and FFM on asthma (Chapter 6).

Whilst expanding MR-Egger to the multivariable setting, it became evident that there may be circumstances where detecting interaction effects between risk factors would be of interest. This observation initiated work on estimating statistical interaction effects in 'factorial' Mendelian randomization (Chapter 5). The methodological framework required to estimate interaction effects between risk factors in Mendelian randomization has not been considered in the literature. However, there has been applied examples on estimating interaction effects between pharmacological interventions, but there remain various unresolved methodological issues relating to this application. Note that this work was not directly relevant to our investigation of the effect of adiposity and body composition on asthma as we did not suspect that there would be statistical interactions in this applied project.

In the next Section, we provide an overview of the dissertation and outline the material presented in each Chapter.

1.7 Structure of the dissertation

The additional assumptions required to estimate causal effects and the most frequently used IV methodology in Mendelian randomization are outlined in Chapter 2. The difficulties of considering binary outcomes in Mendelian randomization are highlighted and discussed in this Chapter. Chapter 2 also contains a literature review on sensitivity analyses in Mendelian randomization that try to identify or account for pleiotropic genetic variants to produce unbiased causal estimates. The review focuses on methods that use summary level data of the genetic associations with the risk factor and with the outcome in Mendelian randomization.

Chapter 3 introduces four robust methods for Mendelian randomization using summary level data. In this Chapter, we assume that heterogeneity among the causal ratio estimates is due to pleiotropic variants. As such, the proposed methods in Chapter 3 remove or downweight the contribution of genetic variants with heterogeneous causal ratio estimates. These methods are compared to other methods in the literature (outlined in Chapter 2) in two applied examples and an extensive simulation study. As highlighted in the acknowledgements, this work is adapted and extended on material uploaded to arXiv by Burgess *et al.* [30] (see Appendix A for a copy of this work). A paper has now been published on this material [31] (see Appendices B and C for a copy of the manuscript and its appendix).

In Chapter 4, we extend the MR-Egger method [29] to the multivariable setting to account for both measured and unmeasured pleiotropy (the 'multivariable MR-Egger method'). Through theoretical arguments, we outline the assumptions required to obtain a consistent causal estimate from the multivariable MR-Egger method. We apply the method to published genetic data, and consider the performance of the method in a simulation study. A paper has already been published on this material [32] (see Appendices D and E for a copy of the manuscript and its appendix).

Chapter 5 presents work on estimating causal interaction effects of risk factors on an outcome in Mendelian randomization analyses. This extension to the Mendelian randomization framework is considered in a simulation study and an applied example. Interaction effects between pharmacological interventions in Mendelian randomization has already been considered in the literature [33–35], and Chapter 5 also addresses some of the unresolved methodological issues relating to this work. A paper has now been published on this material [36] (see Appendices F and G for a copy of the manuscript and its appendix).

Chapter 6 investigates the causal effect of adiposity and body composition on asthma using data from UK Biobank in an extensive one–sample Mendelian randomization

study. A two–sample Mendelian randomization study is also considered by using data from UK Biobank and the GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community) Consortium [37]. Various Mendelian randomization methodology, including the multivariable MR-Egger method developed in Chapter 4, were considered in the studies.

Finally, Chapter 7 discusses the dissertation as a whole, outlines the limitations of the work, and suggests avenues of further research.

Chapter 2

Statistical methods for Mendelian randomization

In this Chapter, we outline the sufficient assumptions required to statistically test for and estimate a causal effect (Sections 2.1 and 2.2), introduce the most commonly used IV methods in Mendelian randomization, and highlight some of the issues of estimating the causal effect with a binary outcome (Section 2.3). We define pleiotropic genetic variants within the context of Mendelian randomization and discuss the impact they may have on the causal estimate (Section 2.4). We also discuss heterogeneous ratio estimates and the exploratory analyses typically performed in Mendelian randomization to detect pleiotropic genetic variants (Section 2.5). Finally, we provide an overview of methods that detect or account for pleiotropic variants using summary level data (Section 2.6).

2.1 Testing for a causal effect

We assume that the genetic variant G satisfies the IV assumptions for a risk factor X and outcome Y. Under the structural assumption (Equation 1.2) in Section 1.4, we can test for a causal effect of X on Y by testing for an association between G and Y. If G is associated with Y, we can infer that the risk factor is causally associated with the outcome [2]. If $Y \perp\!\!\!\perp X|U$ (i.e. there is no arrow between X and Y in Figure 1.1), then the causal effect between X and Y is zero, and $G \perp\!\!\!\perp Y$. However, the converse does not always hold, i.e. if $G \perp\!\!\!\perp Y$ it does not necessarily imply that $Y \perp\!\!\!\perp X|U$; known as the 'non faithfulness' of a DAG [2].

If Y is a continuous variable, we could regress Y against G in a linear regression model to determine whether X has a causal effect on Y. Note that the regression

coefficient will not have a meaningful interpretation and should only be used as a test for a causal effect. We must make additional assumptions to estimate a causal parameter for the effect of X on Y (considered in Section 2.2 below).

2.2 Additional assumptions for a point estimate

Additional modelling assumptions must be made to estimate the causal effect of the risk factor on the outcome. We consider the scenario where we have a continuous risk factor X, continuous outcome Y, unmeasured confounding variables U of the $X - Y$ association, and J independent (not in linkage disequilibrium) genetic variants G_j $(j = 1, \ldots, J)$ that satisfy the IV assumptions.

We assume that for each individual i $(i = 1, \ldots, N_1)$ the risk factor X_i is a linear function of the J genetic variants G_{ij} $(j = 1, \ldots, J)$, the unmeasured confounders U_i of the $X - Y$ association, and the error term ϵ_{Xi}:

$$X_i = \beta_0 + \sum_{j=1}^{J} \beta_{X_j} G_{ij} + \zeta_X U_i + \epsilon_{Xi} \,,$$

where β_{X_j} is the effect of the j^{th} genetic variant on X, and ζ_X is the effect of the unmeasured confounders U on X. G_{ij} is the number of minor alleles at the j^{th} genetic variant for the i^{th} individual, and can take the value 0, 1 or 2. The J genetic associations with the risk factor β_{X_j} $(j = 1, \ldots, J)$ can be estimated by regressing the risk factor against each of the genetic variants in linear regression models, where it is assumed that the minor allele has an additive effect on X. We also assume that for each individual i $(i = 1, \ldots, N_2)$ the outcome is a linear function of the risk factor X_i, the unmeasured confounders U_i of the $X - Y$ association, and the error term ϵ_{Yi}:

$$Y_i = \theta_0 + \theta X_i + \zeta_Y U_i + \epsilon_{Yi} \,, \tag{2.1}$$

where ζ_Y is the effect of the unmeasured confounders U on Y, and θ represents the causal parameter of interest. Under the structural assumption (Equation 1.2), we assume that Equation 2.1 is valid when X is intervened on or observed. Under these model assumptions, the causal parameter is given by:

$$\theta = \frac{\text{cov}(Y, G_j)}{\text{cov}(X, G_j)} \tag{2.2}$$

and this can be estimated by the Wald (ratio) estimator defined in Section 2.3.1. Note that the J genetic associations with the outcome β_{Y_j} $(j = 1, \ldots, J)$ can be estimated by regressing the outcome against each of the genetic variants in linear regression models.

Figure 2.1 contains a DAG of G_j, X, U and Y, where G_j satisfies the IV assumptions. As highlighted in Section 1.5.1, the arrow between G_j and X does not have to be causal, but G_j should be in linkage disequilibrium with the genetic variant that has a causal effect on X. In Figure 2.1, we have included the parameters defined in the model assumptions, i.e. θ represents the causal parameter of interest as defined in Equation 2.2. Strictly speaking, DAGs should provide a non-parametric representation of the relationships between a set of variables, but throughout this dissertation we include the parameters considered in the model assumptions for ease of interpretation.

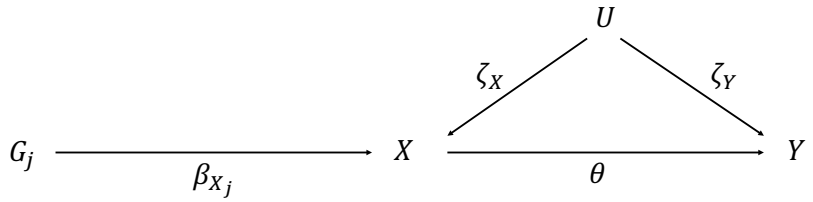


Fig. 2.1 Directed acyclic graph illustrating the Mendelian randomization assumptions for the J genetic variants G_j $(j = 1, \ldots, J)$ to investigate the causal effect of a continuous risk factor X on a continuous outcome Y. The genetic effect of G_j on X is β_{X_j}, and the causal effect of the risk factor X on the outcome Y is θ. U represents the set of unmeasured variables that confound the association between X and Y with effects ζ_X and ζ_Y.

2.3 Estimating the causal effect

Under the model assumptions defined in Section 2.2, we consider the IV methods that are most frequently used in Mendelian randomization to estimate the causal parameter θ in Equation 2.2: the Wald (ratio) estimator that typically uses summary level data (Section 2.3.1) [2]; and two stage-least squares (TSLS) regression that uses individual level data (Section 2.3.1) [38]. Since this dissertation is primarily interested in Mendelian randomization methods that use summary level data, the Wald (ratio) estimator is discussed in detail. Although not considered here, methods based on limited information maximum likelihood [39], generalised methods of moments [40, 41], and Bayesian approaches [42, 43] may also be used to estimate the causal effect. Since

we assume that Y is a continuous variable throughout this Chapter, we highlight some of the issues of estimating the causal effect when the outcome is binary (Section 2.3.3).

2.3.1 Wald (ratio) estimator

We assume that we have summary level data on the risk factor and outcome from two independent samples: the genetic association estimates ($\hat{\beta}_{X_j}$ and $\hat{\beta}_{Y_j}$) and their standard errors ($\text{se}(\hat{\beta}_{X_j})$ and $\text{se}(\hat{\beta}_{Y_j})$) for the J genetic variants G_j ($j = 1, \ldots, J$). The causal effect θ of the risk factor X on the outcome Y can be estimated with one genetic variant G_j using the Wald (ratio) method by dividing the genetic association estimate with the outcome by the genetic association estimate with the risk factor:

$$\hat{\theta}_j = \frac{\hat{\beta}_{Y_j}}{\hat{\beta}_{X_j}} . \tag{2.3}$$

The ratio method can also be applied directly to individual level data. For example, if G_j consisted of two subgroups, then the ratio estimator is the average difference in the risk factor between the two subgroups of G_j divided by the average difference in the outcome between the two subgroups of G_j.

An estimate of the causal effect based on all the genetic variants can also be obtained from the weighted average of the J causal ratio estimates:

$$\hat{\theta}_{IVW} = \frac{\sum_{j=1}^{J} w_j \hat{\theta}_j}{\sum_{j=1}^{J} w_j} , \tag{2.4}$$

where w_j is the inverse-variance of the causal ratio estimate $\hat{\theta}_j$ [44]. The pooled estimate in Equation 2.4 is known as the 'inverse-variance weighted' (IVW) method [45]. Under a fixed effect model, where we assume that there is no heterogeneity among the causal ratio estimates [46], the variance of the IVW estimate is given by:

$$\text{var}(\hat{\theta}_{IVW}) = \frac{1}{\sum_{j=1}^{J} w_j} . \tag{2.5}$$

The inverse-variance weights w_j in Equations 2.4 and 2.5 can be approximated from a delta method expansion of the ratio estimate [47]. The first order approximation of w_j from the delta expansion is most commonly used in the IVW estimator [48]:

$$1^{\text{st}} \text{ order approximation of } w_j = \frac{\hat{\beta}_{X_j}^2}{\text{se}(\hat{\beta}_{Y_j})^2} . \tag{2.6}$$

Equation 2.6 assumes that there is no uncertainty in the genetic associations with the risk factor, known as the NO Measurement Error (NOME) assumption [49]. The NOME assumption will only be satisfied if N_1 is infinite. Since summary level data is obtained from GWASs and consortia with very large sample sizes, the NOME assumption may be considered reasonable.

The causal effect of the risk factor on the outcome can also be estimated using a weighted linear regression of the genetic association estimates with the risk factor ($\hat{\beta}_{X_j}$) and the genetic association estimates with the outcome ($\hat{\beta}_{Y_j}$) [45], with the inverse-variance as weights ($\text{se}(\hat{\beta}_{Y_j})^{-2}$):

$$\hat{\beta}_{Y_j} = \theta_{IVW}\hat{\beta}_{X_j} + \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0, \phi^2\,\text{se}(\hat{\beta}_{Y_j})^2)\,, \tag{2.7}$$

where ϵ_j represents the error term, ϕ represents the residual standard error, and the intercept term is set to zero under the IV2 and IV3 assumptions. To obtain the same variance as the IVW estimate in Equation 2.5, the residual standard error in the weighted linear regression model in Equation 2.7 must be set to one. By fixing ϕ to one, Equation 2.7 is equivalent to performing a fixed-effect meta-analysis of the J causal ratio estimates $\hat{\theta}_j$ $(j = 1, \ldots, J)$ [50].

If heterogeneity among the ratio estimates is suspected, then a multiplicative random-effects model may be preferred to a fixed-effect model. Although the point estimates from the fixed- and random-effect models will be the same, the standard error of the causal estimate from the multiplicative random-effects model will be larger if there is heterogeneity among the ratio estimates. The variance of the IVW estimator under a multiplicative random-effect model with first order weights (Equation 2.6) is given by:

$$\text{var}(\hat{\theta}_{IVW}) = \frac{\hat{\phi}^2}{\sum_{j=1}^{J}\hat{\beta}_{X_j}^2\,\text{se}(\hat{\beta}_{Y_j})^{-2}}\,,$$

where $\hat{\phi}$ is the estimate of the residual standard error. If $\hat{\phi} > 1$, then this suggests that there is over-dispersion in the ratio estimates [50]. Note that it is not biologically plausible for the causal ratio estimates to be under-dispersed ($\hat{\phi} < 1$) if the genetic variants are independent (not in linkage disequilibrium) [46]. $\hat{\phi}$ is not allowed to be lower than one to ensure that the causal estimate from the multiplicative random-effect model is never more precise than the estimate from the fixed-effect model.

Instead of using multiplicative random-effects, an additive random-effects model could be used (not considered throughout this dissertation). This would be equivalent

to performing an additive random-effects meta-analysis of the J causal ratio estimates $\hat{\theta}_j$ $(j = 1, \ldots, J)$ [50]. The estimates and standard errors from the fixed-effects and additive random-effects models will differ if there is heterogeneity among the J causal ratio estimates $\hat{\theta}_j$ $(j = 1, \ldots, J)$. However, additive random-effects are rarely used in Mendelian randomization, with multiplicative random-effects generally being used when heterogeneity among the ratio estimates is suspected. This preference may be due to the fixed-effects and multiplicative random-effects models estimating the same point estimate. Additionally, Bowden *et al.* [51] have cautioned against the use of additive random-effects as weak instruments may be given too much weight under certain scenarios, resulting in more biased estimates of the causal effect under the additive random-effects model than the fixed-effect model.

2.3.2 Two-stage least squares regression

If there is individual level data on the risk factor, outcome, and genetic variants, then the casual effect θ can be estimated using two-stage least squares (TSLS) regression [38] in a one–sample Mendelian randomization study. Under TSLS regression, θ is estimated from the two linear regression models: 1) the regression of the risk factor X against the genetic variants $\boldsymbol{G}$; and 2) the regression of the outcome Y against the predicted values of the risk factor $\hat{X}$ from 1). The coefficient of $\hat{X}$ in the second stage regression model is the TSLS estimate of the causal effect θ. If TSLS is performed manually, the uncertainty in the first stage regression will not have been accounted for, and the causal estimate will be too precise. As such, TSLS regression software should be used to obtain accurate standard errors of the causal estimate. The estimate from the IVW method will be asymptotically equivalent to the estimate from the TSLS method if the genetic variants are uncorrelated [52].

2.3.3 Binary outcomes

Throughout this Section, we only consider linear additive models where the risk factor X and outcome Y are continuous variables. It is likely that the outcome of interest will be binary in an epidemiological study, and the causal odds ratio will be the preferred measure of association. Odds ratios are a non-collapsible measure of association, meaning that if the odds ratio takes a constant value across the strata of a covariate, the value obtained from the marginal analysis may not be equal to this constant value [53]. Whilst the numerator in Equation 2.3 could be replaced with the estimate of the log odds ratio of the j^{th} genetic variant with the outcome, and the

second stage regression model in TSLS be replaced with logistic regression, due to the non-collapsibility of the odds ratio, these estimators will generally not produce consistent estimates of the causal odds ratio [54]. If the outcome is rare, then these versions of the ratio estimator and TSLS regression will approximate the causal risk ratio [55]. Hence, the IVW method and TSLS regression will only provide approximate measures of the causal odds ratio for the effect of a continuous risk factor X on a binary outcome Y. TSLS with a logistic regression in the second stage will, however, provide a valid test for the null hypothesis of no causal association [56].

The Mendelian randomization study on the effect of adiposity and body composition on asthma, considered in Chapter 6, has a binary outcome measurement. We note that the methods in Chapters 3 to 5 have been primarily developed under the assumption that the outcome is a continuous variable. We do investigate in a simulation study whether the methods proposed in Chapter 3 provide a good approximation of the causal effect when the outcome is binary rather than continuous. The limitations of considering a continuous outcome variable in Chapters 3 to 5 are highlighted and discussed in each Chapter.

2.4 Pleiotropic genetic variants

In this Section, we define a pleiotropic genetic variant within the context of Mendelian randomization. We consider pleiotropic effects and the impact they may have on the causal estimate from the IVW method.

In a Mendelian randomization study, a pleiotropic variant is defined as a genetic variant G_j that is associated with multiple traits, including the risk factor X of interest (Figure 2.2). If a pleiotropic genetic variant is associated with traits that mediate the relationship between the risk factor and the outcome, and the genetic variant has no direct effect on these mediators, i.e. the mediators lie on the same causal pathway as the genetic variant and risk factor, then the IV3 assumption will not be violated (referred to as 'vertical pleiotropy'). For example, we may suspect that age at menarche has an effect on body mass index (BMI), and BMI has an effect on breast cancer. If the pleiotropic genetic variant used as an IV for age at menarche has no direct effect on BMI, then the IV3 assumption will not be violated. However, if the variant is associated with multiple traits on causal pathways that are independent of the risk factor, then the IV2 or IV3 assumptions will be violated (referred to as 'horizontal pleiotropy').

Figure 2.2 contains examples of horizontal pleiotropy, where δ_j represents the effect of the genetic variant on the unmeasured confounders, and α_j represents the direct effect of G_j on the outcome. The IV2 assumption would be violated if $\delta_j \neq 0$, and the IV3 assumption would be violated if $\alpha_j \neq 0$. Since it is not possible to distinguish between vertical and horizontal pleiotropy, and the biological mechanisms between the genetic variants and the risk factor are rarely understood in a Mendelian randomization study [57], the inclusion of pleiotropic genetic variants is a real concern.

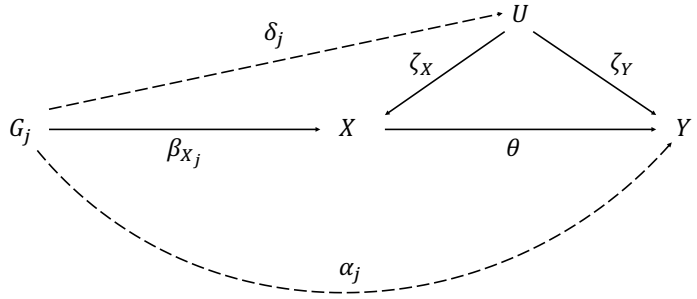


Fig. 2.2 Directed acyclic graph illustrating Mendelian randomization assumptions with potential violation of IV2 or IV3 by a pleiotropic effect indicated by two dotted lines. The genetic effect of G_j on X is β_{X_j}, the genetic effect of G_j on U is represented by δ_j (representing the potential violation of the IV2 assumption), the direct effect of G_j on Y is α_j (representing the potential violation of the IV3 assumption), and the causal effect of the risk factor X on the outcome Y is θ. U represents the set of unmeasured variables that confound the association between X and Y with effects ζ_X and ζ_Y.

Throughout this dissertation we only consider pleiotropy with respect to the violation of the IV3 assumption: a pleiotropic variant is associated with the outcome Y via a causal pathway that is independent of the risk factor X and the set of unmeasured confounding variables U ($\alpha_j \neq 0$ in Figure 2.2). Under this definition of pleiotropy, and assuming that all of the associations are linear with no effect modification, we can express the genetic association with the outcome as a linear combination of the direct effect and the indirect effect via the risk factor:

$$\beta_{Yj} = \alpha_j + \theta\beta_{X_j} .$$

The genetic variant G_j is pleiotropic and violates the IV3 assumption if $\alpha_j \neq 0$. If the genetic variant G_j is pleiotropic, then the ratio estimand for the causal effect of X on

Y for the j^{th} genetic variant would be:

$$\theta_j = \theta + \frac{\alpha_j}{\beta_{X_j}},$$

where α_j/β_{X_j} is the non-zero bias term. If $\boldsymbol{G}$ contains multiple pleiotropic variants whose average direct effect is zero (referred to as 'balanced pleiotropy'), and the direct effects are independent of the genetic associations with the risk factor, then the IVW method will produce consistent causal estimates [51]. If the average direct effect differs from zero (referred to as 'directional pleiotropy'), then the IVW estimate will be biased.

2.5 Heterogeneity and pleiotropic variants

There should be little heterogeneity in the ratio estimates $\boldsymbol{\theta}$ if all of the genetic variants $\boldsymbol{G}$ are valid IVs. Heterogeneity among the causal ratio estimates may provide evidence that some of the genetic variants are pleiotropic. As such, the validity of the IV assumptions for genetic variants with outlying or heterogeneous ratio estimates should be considered in more detail.

In this Section, we outline the exploratory analyses typically used in Mendelian randomization to assess heterogeneity among the ratio estimates: plots of the genetic associations; and a formal test for heterogeneity. Finally, we discuss the limitation of assuming that genetic variants with heterogeneous ratio estimates are pleiotropic.

2.5.1 Plots of the genetic associations

Scatter plots of the genetic association estimates with the outcome $\hat{\boldsymbol{\beta}}_{\boldsymbol{Y}}$ against the genetic association estimates with the risk factor $\hat{\boldsymbol{\beta}}_{\boldsymbol{X}}$ are frequently used in Mendelian randomization to inspect summary level data (used throughout this dissertation) [48]. Figure 2.3 contains an example scatter plot (considered in the applied example in Chapter 3) of the genetic associations and 95% confidence intervals with Alzheimer's disease (outcome of interest) against the genetic associations and 95% confidence intervals with low-density lipoprotein (risk factor of interest) for 75 genetic variants. Each data point represents a single genetic variant, and the gradient of the line connecting the data point to the origin represents the ratio estimate $\hat{\theta}_j$ (dotted line in Figure 2.3 is provided as an example). The weighted average of the ratio estimates from the IVW method is usually displayed in these scatter plots to allow for comparisons between the J ratio estimates and the IVW estimate (solid line in Figure 2.3).

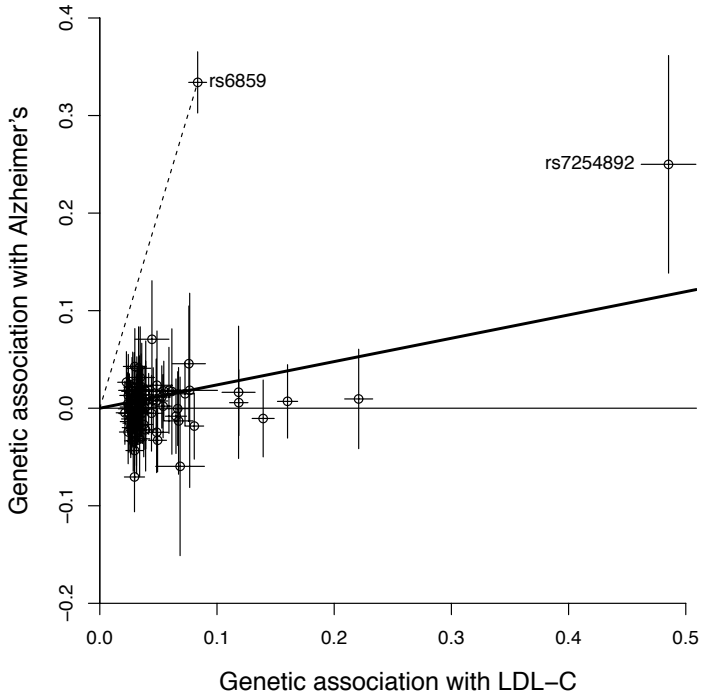


Fig. 2.3 Scatter plot of the estimated genetic associations and 95% confidence intervals with Alzheimer's disease (log odds ratios) against the estimated genetic associations and 95% confidence intervals with low-density lipoprotein (LDL-C, standard deviation units) for 75 genetic variants. The dotted line represents the ratio estimate for the rs6859 genetic variant, and the solid line is the inverse-variance weighted estimate of the effect of LDL-C on Alzheimer's disease. See Section 3.4 for more detail on the dataset used, and the analyses performed.

The scatter plots of $\hat{\boldsymbol{\beta}}_{\boldsymbol{Y}}$ against $\hat{\boldsymbol{\beta}}_{\boldsymbol{X}}$ may highlight heterogeneity among the ratio estimates. We should observe a dose-response relationship in the scatter plot if all of the genetic variants are valid IVs. A pleiotropic variant may appear as an outlier if the majority of the genetic variants are valid IVs. In Figure 2.3, we may suspect that the genetic variants rs6859 and rs7254892 are pleiotropic as they are outliers to the main group of genetic variants.

Funnel plots used in the meta-analysis literature can be used in Mendelian randomization to detect directional pleiotropy by plotting the precision of the ratio estimates ($\sqrt{\boldsymbol{w}}$) against $\hat{\boldsymbol{\theta}}$ [58, 29]. If the genetic variants are valid IVs, the plot should be a symmetric funnel with the precise ratio estimates having less variability [59]. Asymmetry in the funnel plot suggests that there is heterogeneity among the ratio estimates, and the average direct effect of the genetic variants may be non-zero (directional pleiotropy).

Bowden *et al.* [48] have adapted the Galbraith Radial plot used in the meta-analysis literature [60] as a replacement to the scatter plot in Figure 2.3. The Radial plot

provides a more transparent indication of the contribution each genetic variant makes to the IVW estimate. The authors have also proposed a Radial funnel plot [48].

2.5.2 Test for heterogeneity

Cochran's Q statistic is used in meta-analysis to test for heterogeneity [61]. Del Greco *et al.* [62] first introduced the idea of using the Q statistic to test for heterogeneity among the ratio estimates in Mendelian randomization:

$$Q = \sum_j w_j(\hat{\theta}_j - \hat{\theta}_{IVW})^2 \,, \tag{2.8}$$

where Q has an approximate χ^2_{J-1} distribution under the null hypothesis that all J genetic variants satisfy the IV assumptions [62]. If the Q statistic provides evidence of heterogeneity among the ratio estimates then a multiplicative random-effects model should be used in the main analysis (as discussed in Section 2.3.1). A statistically significant Q statistic may suggest that some of the genetic variants in $\boldsymbol{G}$ are pleiotropic. The Type I error rate of detecting heterogeneity among the ratio estimates may be inflated if the first order approximations of w_j (Equation 2.6) are used in Equation 2.8 [63]. This limitation has been addressed by Bowden *et al.* [63] who have proposed a modified weight for w_j.

2.5.3 Limitations

Although the exploratory analyses described above are useful tools in detecting heterogeneity among the ratio estimates, their utility may be limited. We have assumed that heterogeneity among the causal ratio estimates is due to pleiotropic genetic variants. Whilst this assumption may be correct, there are other reasons why the ratio estimates may be heterogeneous, including population stratification, systematic measurement errors in the risk factor and outcome, and non-linearity.

Scatter plots of the genetic associations are easy to interpret and provide a good overview of the data in a Mendelian randomization study. However, the plots should not be used uncritically as a tool to exclude genetic variants. Genetic variants that appear as 'obvious' outliers in the scatter plot of the genetic associations may be perfectly valid IVs. Instead, the plots should be used as a tool to identify genetic variants that require additional investigation into the plausibility of the IV assumptions. Unlike the scatter plots, the Q statistic quantifies whether there is heterogeneity among the ratio estimates. Whilst this can be helpful in determining whether fixed- or random-effects

should be used in the IVW model, it does not provide any information on the source of the heterogeneity, and in particular, whether it is a result of balanced or directional pleiotropy.

In the next Section, we discuss methods that systematically account for pleiotropic variants in Mendelian randomization analyses.

2.6 Sensitivity analyses

In this Section, we review the methods in the IV literature that detect or account for pleiotropic genetic variants using summary level data. These methods can be divided into two broad categories: methods that downweight or remove genetic variants (Section 2.6.1); and methods that estimate consistent causal effects in the presence of pleiotropic genetic variants without downweighting their contribution to the causal estimate. For the second category, we discuss two methods: one method that relaxes the IV3 assumption and introduces a weaker assumption (MR-Egger method in Section 2.6.2); and another method that extends the IV assumptions to the multivariable setting ('multivariable Mendelian randomization' in Section 2.6.3). We discuss these two methods in detail as they are instrumental to Chapters 4 and 5.

Although this review focuses on methods that use summary level data, we acknowledge that there are many methods in the IV literature that detect or account for pleiotropic variants using individual level data [64–69].

2.6.1 Downweighting or removing genetic variants

In this Section, we consider methods that try to identify pleiotropic genetic variants to downweight their contribution to the causal estimate, or exclude them from the analysis. These methods assume that there is a subgroup of genetic variants that satisfy the IV assumptions. Note that genetic variants may be assigned a weight of zero but are not explicitly removed from the dataset.

Downweighting genetic variants

The simple median estimator is the median of the J ratio estimates $\hat{\theta}_j$ $(j = 1, \ldots, J)$ [52]. It will produce consistent causal estimates if at least 50% of the genetic variants are valid IVs, known as the '50% rule' or 'majority rule' [70]. The standard error of the simple median estimate is obtained through bootstrapping methods. If there is variability in the precision of the ratio estimates then the efficiency of the median

estimator can be improved by using the inverse-variance weights, known as the 'weighted median estimator' [52]. The weighted median estimate is the 50^{th} percentile of the inverse-variance weighted empirical distribution of $\hat{\boldsymbol{\theta}}$. The estimates will be consistent if 50% or more of the weights from valid IVs contribute to the weighted median estimate.

If more than 50% of the genetic variants are invalid IVs, then the simple median estimator will be biased. To overcome this limitation, Hartwig *et al.* [71] have introduced a mode-based estimator (MBE) that produces asymptotically consistent causal estimates when more than 50% of the variants are invalid IVs. The estimates from the MBE will be consistent if the valid IVs make up the largest subset of homogeneous ratio estimates, known as the 'zero modal pleiotropy assumption' (ZEMPA). The estimate from the MBE is the mode of the smoothed empirical density function of the J ratio estimates:

$$f(x) = \frac{1}{h\sqrt{2\pi}} \sum_{j=1}^{J} w_{MBE_j} \exp\Big[-\frac{1}{2}\Big(\frac{x - \hat{\theta}_j}{h}\Big)^2 \Big],$$

where the causal estimate x is the value that maximizes $f(x)$, h is the smoothing bandwidth, and $\boldsymbol{w_{MBE}}$ are the weights [71]. The ratio estimates can either have an equal contribution to x under the 'simple MBE', or standardized inverse-variance weights can be used under the 'weighted MBE'. The value of h must be specified by the user, with larger values producing more precise estimates. Results from the MBE can be highly sensitive to the values of h [72]. Simulation studies have shown that the MBE may be less efficient than the weighted median estimator [71].

Burgess *et al.* [72] have developed a heterogeneity-penalized model-averaging method that is based on the ZEMPA assumption. The overall causal estimate from this method is the mode of the mixture distribution of the estimates obtained from each possible subset of genetic variants (excluding subsets with 0 or 1 genetic variants). Larger subsets of genetic variants are given a greater weight in the mixture distribution unless their ratio estimates are heterogeneous, in which case, the contribution of the subset to the mixture model is dramatically reduced. Unlike the method proposed by Hartwig *et al.* [71], a bandwidth does not need to be specified by the user, and the standard errors are obtained without using bootstrapping methods [72].

Although the median- or mode-based estimators do not explicitly remove any of the genetic variants from the analysis, some of the genetic variants will have no direct contribution to the overall causal estimate. As such, these methods may be less efficient than methods that explicitly exclude genetic variants from the analysis (considered below).

Removing genetic variants

The Mendelian randomization pleiotropy residual sum and outlier (MR-PRESSO) method has three main functions using summary level data: 1) to test for directional pleiotropy (MR-PRESSO global test); 2) to identify outlying variants that may be pleiotropic (MR-PRESSO outlier test); and 3) to empirically test the difference between the IVW estimates using the full and reduced sets of genetic variants [73]. The MR-PRESSO global test is performed by calculating the global observed residual sum of squares (RSS_{obs}):

$$\text{RSS}_{obs} = \sum_{j=1}^{J} \text{RSS}_{obs_j} = \sum_{j=1}^{J} (\hat{\beta}_{Y_j} - \hat{\theta}_{IVW_{-j}} \hat{\beta}_{X_j})^2 \, \text{se}(\hat{\beta}_{Yj})^{-2} \,,$$

where $\hat{\theta}_{IVW_{-j}}$ is the estimate from the IVW model when the j^{th} genetic variant has been removed. RSS_{obs} is compared against a simulated expected distribution of the residual sum of squares under the null hypothesis of no pleiotropy. The expected RSS is given by:

$$\text{RSS}_{exp} = \sum_{j=1}^{J} \text{RSS}_{exp_j} = \sum_{j=1}^{J} (\hat{\beta}'_{Y_j} - \hat{\theta}_{IVW_{-j}} \hat{\beta}'_{X_j})^2 \, \text{se}(\hat{\beta}_{Yj})^{-2} \,,$$

where the genetic associations $\hat{\beta}'_{X_j}$ and $\hat{\beta}'_{Y_j}$ are simulated from the normal distributions:

$$\hat{\beta}'_{X_j} \sim \mathcal{N}(\hat{\beta}_{X_j}, \text{se}(\hat{\beta}_{X_j})^2) \text{ and } \hat{\beta}'_{Y_j} \sim \mathcal{N}(\hat{\theta}_{IVW_{-j}} \hat{\beta}_{X_j}, \text{se}(\hat{\beta}_{Y_j})^2) \,.$$

RSS_{exp} is generated N times to obtain a distribution of N expected residual sum of squares. An empirical p-value for the global test of directional pleiotropy is calculated as the proportion of times the N expected residual sum of squares is greater than RSS_{obs}. Verbanck *et al.* [73] recommend that $N \geq 1,000$ to ensure there is adequate precision of the p-value. Note that the individual residual sum of squares RSS_{obs_j} can be used to identify pleiotropic genetic variants, allowing for the 'corrected' causal estimate to be obtained from the IVW method based on the reduced set of genetic variants.

The global and individual tests for direct effects (GLIDE) method has also been proposed to detect pleiotropy in summary level data [74]. Like the MR-PRESSO method, GLIDE tests for global pleiotropy among the genetic variants and tries to identify and exclude genetic variants that may be pleiotropic. This method uses Q-Q plots and permutation procedures to identify pleiotropic genetic variants. Rather than

considering a continuous outcome, the GLIDE method expresses the causal effect in terms of the relative risk.

Other methods have been proposed that identify and remove pleiotropic variants, including the heterogeneity in dependent instruments (HEIDI) method that was introduced under the summary data-based Mendelian randomization (SMR) method [75], and was developed further under the generalized SMR framework (GSMR).

2.6.2 MR-Egger

The MR-Egger (Mendelian randomization-Egger) method [29] was adapted from Egger regression, a tool used in meta-analysis to detect small study bias [76]. MR-Egger can be used to estimate consistent causal effects in the presence of pleiotropic genetic variants, and to test the validity of the IV assumptions. The MR-Egger method replaces the IV3 condition with the untestable assumption that the genetic associations with the risk factor are independent of the direct effects of the genetic variants on the outcome ($\boldsymbol{\beta_X} \perp\!\!\!\perp \boldsymbol{\alpha}$), known as the InSIDE (instrument strength independent of direct effect) assumption. Like the IVW method, MR-Egger also assumes that the NOME assumption is satisfied.

The MR-Egger method fits the weighted linear regression of the genetic association estimates with the risk factor ($\hat{\beta}_{X_j}$) and the genetic association estimates with the outcome ($\hat{\beta}_{Y_j}$) [45], with the inverse-variance as weights ($\text{se}(\hat{\beta}_{Y_j})^{-2}$) and the intercept unrestrained:

$$\hat{\beta}_{Y_j} = \theta_{0E} + \theta_E \hat{\beta}_{X_j} + \epsilon_{E_j} \quad \epsilon_{E_j} \sim \mathcal{N}(0, \phi_E^2 \, \text{se}(\hat{\beta}_{Y_j})^2) \,, \tag{2.9}$$

where θ_{0E} is the intercept term, θ_E is the MR-Egger causal effect, ϵ_{E_j} is the error term, and ϕ_E represents the residual standard error under the MR-Egger model. The interpretation of the estimates $\hat{\theta}_{0E}$ and $\hat{\theta}_E$ from the MR-Egger method are discussed alongside the InSIDE and NOME assumptions in the subsections below.

$\hat{\theta}_E$ and the InSIDE assumption

We initially assume that there is no estimation (or measurement) error in the genetic associations with the risk factor, i.e. the NOME assumption is satisfied. If there are no pleiotropic effects ($\boldsymbol{\alpha} = 0$), the MR-Egger estimate $\hat{\theta}_E$ should be asymptotically equivalent to the IVW estimate $\hat{\theta}_{IVW}$ (Equation 2.7). The InSIDE assumption must be satisfied for $\hat{\theta}_E$ to be a consistent estimate of the causal effect θ in the presence of balanced or directional pleiotropy [29, 64]. If the InSIDE assumption is satisfied, then

the weighted covariance of $\boldsymbol{\beta_X}$ and $\boldsymbol{\alpha}$ ($\text{cov}_w(\boldsymbol{\alpha}, \boldsymbol{\beta_X})$) will tend to zero as the number of genetic variants J tends to infinity. The estimate of θ from MR-Egger is:

$$\hat{\theta}_E = \frac{\text{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{Y}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X}})}{\text{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X}})} \xrightarrow{N\to\infty} \frac{\text{cov}_w(\boldsymbol{\beta_Y}, \boldsymbol{\beta_X})}{\text{var}_w(\boldsymbol{\beta_X})} = \theta + \frac{\text{cov}_w(\boldsymbol{\alpha}, \boldsymbol{\beta_X})}{\text{var}_w(\boldsymbol{\beta_X})}, \tag{2.10}$$

which is equal to θ if the InSIDE assumption is satisfied, where cov_w and var_w represent the weighted covariance and weighted variance using the inverse-variance as weights $\text{se}(\hat{\beta}_{Y_j})^{-2}$:

$$\begin{aligned}
\text{cov}_w(\boldsymbol{\alpha}, \boldsymbol{\beta_X}) &= \frac{\sum_j (\alpha_j - \bar{\alpha}_w)(\beta_{X_j} - \bar{\beta}_{X_w})\,\text{se}(\hat{\beta}_{Y_j})^{-2}}{\sum_j \text{se}(\hat{\beta}_{Y_j})^{-2}}, \\
\text{var}_w(\boldsymbol{\beta_X}) &= \frac{\sum_j (\beta_{X_j} - \bar{\beta}_{X_w})^2\,\text{se}(\hat{\beta}_{Y_j})^{-2}}{\sum_j \text{se}(\hat{\beta}_{Y_j})^{-2}}, \\
\bar{\alpha}_w &= \frac{\sum_j \alpha_j\,\text{se}(\hat{\beta}_{Y_j})^{-2}}{\sum_j \text{se}(\hat{\beta}_{Y_j})^{-2}}, \\
\bar{\beta}_{X_w} &= \frac{\sum_j \beta_{X_j}\,\text{se}(\hat{\beta}_{Y_j})^{-2}}{\sum_j \text{se}(\hat{\beta}_{Y_j})^{-2}}.
\end{aligned}$$

If the InSIDE assumption is violated, and there is balanced or directional pleiotropy, then the estimate of θ from MR-Egger will be biased due to the non-zero bias term in Equation 2.10.

It seems more plausible that the InSIDE assumption will hold if the pleiotropic effects are independent of the unmeasured variables U that confound the $X - Y$ association ($\boldsymbol{\delta} = 0$ in Figure 2.2). If the pleiotropic variants do lie on the same causal pathway as the unmeasured confounders U of the $X - Y$ association ($\boldsymbol{\delta} \neq 0$), then it is difficult to conceive how the InSIDE assumption would be satisfied as the strength of the genetic associations with X will depend on the strength of the pleiotropic effects via U. As noted in Section 2.4, throughout this dissertation we assume that a pleiotropic variant is associated with the outcome Y via a causal pathway that is independent of the risk factor X and the set of unmeasured confounding variables U ($\alpha_j \neq 0$ in Figure 2.2).

In terms of the standard error of $\hat{\theta}_E$, estimating the intercept term in Equation 2.9 will result in less precise estimates of the causal effect from the MR-Egger method compared to the IVW method. Since the MR-Egger method allows for the possibility of the genetic variants to be pleiotropic, applying a fixed-effects model to Equation 2.9 would not be logical. A multiplicative random-effects model, where the residual standard

error ϕ_E is estimated, is therefore applied to the MR-Egger method throughout this dissertation.

Orientation of the genetic variants

In this dissertation, we assume that G_j $(j = 1, \ldots, J)$ can take the value 0, 1 or 2, representing the number of risk-increasing or risk-decreasing alleles of a bi-allelic genetic variant. The interpretation of the genetic associations with the risk factor or outcome will depend on whether G_j relates to the risk-increasing allele or risk-decreasing allele. If G_j has been orientated with respect to the risk-increasing allele for X, then the genetic association with the risk factor represents the average change in X per additional copy of the risk factor-increasing allele. Since the intercept term in Equation 2.4 is fixed at zero, the orientation of the genetic variants has no affect on the estimate of the causal effect θ from the IVW method.

The orientation of the genetic variants will affect the estimates $\hat{\theta}_{0E}$ and $\hat{\theta}_E$ from the MR-Egger method as the orientation will determine the definition of the pleiotropic effect. Hence, the orientation of the genetic variants will also affect the defintion of the InSIDE assumption. Bowden *et al.* [29] therefore suggest that the genetic variants be orientated to ensure the direction of the genetic associations with the risk factor are either positive for all variants or negative for all variants.

$\hat{\theta}_{0E}$ and the MR-Egger intercept test

If we assume that the genetic variants G_j $(j = 1, \ldots, J)$ are orientated with respect to the risk factor-increasing alleles, and the InSIDE assumption is satisfied, then the estimate of the intercept term $\hat{\theta}_{0E}$ can be interpreted as the average direct effect of the J genetic variants with respect to the risk factor-increasing alleles [52]. The InSIDE assumption will be satisfied if $\boldsymbol{\beta_X} \perp\!\!\!\perp \boldsymbol{\alpha}$ when the genetic variants are orientated with respect to the risk factor-increasing alleles. If there is balanced pleiotropy, and the InSIDE assumption is valid, then the intercept term should tend to zero as the sample size increases. If the intercept term differs from zero, then either the InSIDE assumption is violated, or there is directional pleiotropy, or both conditions are violated. Testing the intercept term in Equation 2.9 is a way of assessing the validity of the IV assumptions, and is known as the 'MR-Egger intercept test' [46].

Violation of the NOME assumption

We now consider the impact the violation of the NOME assumption has on the MR-Egger method. First consider the weighted variance of the genetic associations with the risk factor:

$$\mathrm{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X}}) = \mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}}) + s_w^2\,,$$

where s_w^2 is the weighted average of the variability in $\hat{\boldsymbol{\beta}}_{\boldsymbol{X}}$ explained by estimation (or measurement) error. If the NOME assumption is satisfied, there is no uncertainty in the genetic associations with the risk factor, and s_w^2 is equal to zero.

If the InSIDE assumption is satisfied, then the expected value of the MR-Egger estimate can be expresssed as [49]:

$$\mathbb{E}[\hat{\theta}_E] = \mathbb{E}\left[\frac{\mathrm{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{Y}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X}})}{\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}})}\frac{\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}})}{\mathrm{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X}})}\right] \approx \theta\frac{\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}})}{\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}}) + s_w^2}\,, \tag{2.11}$$

and the MR-Egger method will produce a consistent estimate of the casual effect if the NOME assumption is satisfied. From Equation 2.11, the MR-Egger estimate will be attenuated towards zero if the NOME assumption is violated ($s_w^2 \neq 0$). Violation of the NOME assumption will also lead to an increased Type I error rate for the MR-Egger intercept test [49].

The extent to which the MR-Egger estimate $\hat{\theta}_E$ is attenuated is dependent upon $\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}})$ and s_w^2. If there is a lot of variability in $\boldsymbol{\beta}_{\boldsymbol{X}}$, and little estimation error, then the attenuation of the MR-Egger estimate towards the null will be small. However, if there is little variability in $\boldsymbol{\beta}_{\boldsymbol{X}}$ relative to the estimation error, then the attenuation of the MR-Egger estimate will be more severe.

To account for the violation of the NOME assumption in Equation 2.11, we require an estimate of $\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}})/\,\mathrm{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X}})$. Bowden *et al.* [49] have shown that $\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}})/\,\mathrm{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X}})$ can be estimated through an adapted version of the I^2 statistic used in the meta-analysis literature to assess heterogeneity:

$$I^2 = \frac{(Q_{GX} - (J-1))}{Q_{GX}}\,, \tag{2.12}$$

where Q_{GX} is Cochran's Q statistic for the genetic associations with the risk factor:

$$Q_{GX} = \sum_{j=1}^{J}\frac{(\hat{\beta}_{X_j}\,\mathrm{se}(\hat{\beta}_{Y_j})^{-1} - \bar{\hat{\beta}}_X)^2}{\mathrm{se}(\hat{\beta}_{X_j})^2\,\mathrm{se}(\hat{\beta}_{Y_j})^{-2}}\,,$$

and $\bar{\hat{\beta}}_X$ is the mean of the genetic associations with the risk factor weighted by $\text{se}(\hat{\beta}_{X_j})^{-2}$. The I^2 statistic will lie between 0 and 1, with smaller values corresponding to more biased MR-Egger estimates. If the I^2 statistic is close to 1, then there should be little or no attenuation of the causal estimate from the MR-Egger method.

Since Bowden *et al.* [49] obtained unstable results when the MR-Egger estimate $\hat{\theta}_E$ was divided by I^2, the authors suggest that a simulation extrapolation (SIMEX) method be used to adjust for the violation of the NOME assumption when $I^2 < 0.9$. Under the SIMEX approach, estimates of the genetic associations with the risk factor $\hat{\beta}^{\lambda}_{X_j}$ $(j = 1, \ldots, J)$ are simulated from:

$$\hat{\beta}^{\lambda}_{X_j} \sim \mathcal{N}(\hat{\beta}_{X_j}, \lambda\, \text{se}(\hat{\beta}^{\lambda}_{X_j})^2)\,,$$

where $\hat{\beta}_{X_j}$ and $se(\hat{\beta}^{\lambda}_{X_j})$ are the observed data, and $var(\hat{\boldsymbol{\beta}}^{\boldsymbol{\lambda}}_{\boldsymbol{X}}) = (1+\lambda)se(\hat{\boldsymbol{\beta}}^{\boldsymbol{\lambda}}_{\boldsymbol{X}})^2$. For a given value of $\lambda > 0$, the simulated genetic associations $\hat{\beta}^{\lambda}_{X_j}$ $(j = 1, \ldots, J)$ and the observed genetic associations with the outcome $\hat{\beta}_{Y_j}$ $(j = 1, \ldots, J)$ are used to obtain a MR-Egger estimate of the causal effect. This process is repeated multiple times to obtain an average value of the MR-Egger estimate for a specific value of λ. This whole process is then applied to a range of λ values that increase in small increments. As λ increases, the average MR-Egger estimate will decrease as there will be more attenuation towards zero. The average values for the MR-Egger estimates from the different values of λ are extrapolated to estimate what the MR-Egger estimate may have been if the NOME assumption had been satisfied.

Instrument strength

The strength of the association between the genetic variants and the risk factor for the IVW method is usually assessed through the F-statistic from the regression of the risk factor on the genetic variant(s). Genetic variants are often classified as 'weak' IVs if they have a F-statistic less than 10. Weak IVs will produce asymptotically unbiased causal estimates, but under finite samples they will bias the causal estimate (known as 'weak instrument bias') [77, 78]. For one–sample Mendelian randomization, this bias will be towards the confounded observational association, and for two–sample Mendelian randomization the bias will be towards the null [79].

For MR-Egger, instrument strength should be assessed through the I^2 statistic (Equation 2.12) rather than the F-statistic. An I^2 value close to 1 suggests that the MR-Egger estimate does not suffer from weak instrument bias. If the I^2 statistic is equal to 0.9, then the attenuation of the MR-Egger estimate towards the null will

be approximately 10% of the causal effect θ. To correspond with the classification of 'weak' IVs under the F-statistic for the IVW method, Bowden *et al.* [49] suggest that the SIMEX method be applied when the I^2 is less than 0.9.

2.6.3 Multivariable Mendelian randomization

Multivariable Mendelian randomization assumes that the direct effect $\boldsymbol{\alpha}$ of the genetic variants on the outcome is fully mediated through additional measured risk factors. Rather than replacing the IV3 assumption with a weaker assumption (as done by the MR-Egger method in Section 2.6.2), multivariable Mendelian randomization expands the IV assumptions to allow for the causal effects of multiple risk factors on the outcome to be estimated in the same model. Whilst the MR-Egger method should only be included in the sensitivity analysis of a Mendelian randomization study, multivariable Mendelian randomization can be used as a sensitivity analysis, or as the primary analysis model.

Instrumental variable assumptions

Suppose we have K continuous risk factors X_k $(k = 1, \ldots, K)$, a continuous outcome Y, and K sets of unmeasured confounding variables U_k $(k = 1, \ldots, K)$ of the $\boldsymbol{X} - Y$ associations. The following assumptions must be satisfied in a multivariable Mendelian randomization analysis:

- IV1(M): each genetic variant G_j $(j = 1, \ldots, J)$ is associated with at least one of the K risk factors X_k $(k = 1, \ldots, K)$,
- IV2(M): each risk factor X_k $(k = 1, \ldots, K)$ is associated with at least one of the J genetic variants G_j $(j = 1, \ldots, J)$,
- IV3(M): the variants G_j $(j = 1, \ldots, J)$ are independent of all unmeasured confounders $\boldsymbol{U}$ of each of the risk factor–outcome associations, and
- IV4(M): the variants G_j $(j = 1, \ldots, J)$ are independent of the outcome Y conditional on the risk factors $\boldsymbol{X}$ and confounders $\boldsymbol{U}$ [28].

From the above conditions, each genetic variant G_j $(j = 1, \ldots, J)$ must be associated with at least one of the risk factors X_k $(k = 1, \ldots, K)$, and each risk factor X_k $(k = 1, \ldots, K)$ must be associated with at least one of the genetic variants G_j $(j = 1, \ldots, J)$. Genetic variants that are associated with multiple risk factors can be used in multivariable Mendelian randomization provided that these risk factors are included

in the analysis. There must be as many genetic variants as there are risk factors for the causal effects to be estimated, i.e. $J \geq K$.

Figure 2.4 contains a DAG where the IV assumptions for multivariable Mendelian randomization are satisfied for $K = 3$ risk factors. We assume that for each individual i $(i = 1, \ldots, N_1)$ each risk factor X_{k_i} is a linear function of the J genetic variants G_j $(j = 1, \ldots, J)$, the unmeasured confounders U_{k_i} of the $X_k - Y$ association, and the error term $\epsilon_{X_{k_i}}$:

$$X_{k_i} = \beta_{k0} + \sum_{j=1}^{J} \beta_{X_{k_j}} G_{ij} + U_{k_i} + \epsilon_{X_{k_i}} ,$$

where $\beta_{X_{k_j}}$ is the effect of the j^{th} genetic variant on X_k. G_{ij} is the number of minor alleles at the j^{th} genetic variant for the i^{th} individual, and can take the value 0, 1 or 2. The J genetic associations with the risk factor X_k can be estimated by regressing X_k against each of the genetic variants in linear regression models, where it is assumed that the minor allele has an additive effect on X_k. We also assume that for each individual i $(i = 1, \ldots, N_2)$ the outcome is a linear function of the risk factors X_k $(k = 1, \ldots, K)$, the unmeasured confounders U_{k_i} $(k = 1, \ldots, K)$ of the $\boldsymbol{X} - Y$ associations, and the error term ϵ_{Y_i}. For $K = 3$, we can express Y_i as:

$$Y_i = \theta_0 + \theta_1 X_{1_i} + \theta_2 X_{2_i} + \theta_3 X_{3_i} + U_{1_i} + U_{2_i} + U_{3_i} + \epsilon_{Y_i} .$$

where $\boldsymbol{\theta}$ are the direct effects of the risk factors on the outcome. The J genetic associations with the outcome β_{Y_j} $(j = 1, \ldots, J)$ can be estimated by regressing the outcome against each of the genetic variants in linear regression models.

The aim of a multivariable Mendelian randomization analysis is to estimate the direct effects of the risk factors on the outcome, when conditioned on each other. These estimates can be obtained by using individual level data in one–sample multivariable Mendelian randomization or summary level data in two–sample multivariable Mendelian randomization data as outlined in the subsections below (considered when $K = 3$).

Individual level data

We assume that we have individual level data for the J genetic variants G_j $(j = 1, \ldots, J)$, the three risk factors, and outcome on the same set of participants. Consistent estimates of $\boldsymbol{\theta}$ can be obtained from TSLS regression in a one–sample multivariable Mendelian

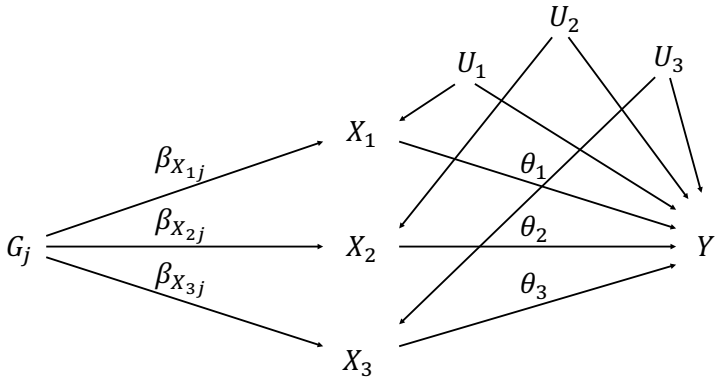


Fig. 2.4 Directed acyclic graph illustrating the multivariable Mendelian randomization assumptions for the J genetic variants G_j $(j = 1, \ldots, J)$ to investigate the causal effect of $K = 3$ continuous risk factors X_k $(j = k, \ldots, K)$ on a continuous outcome Y. The genetic effect of G_j on X_k is $\beta_{X_{kj}}$, and the direct causal effect of the risk factor X_k on the outcome Y is θ_k. U_k represents the set of unmeasured variables that confound the associations between X_k and Y.

randomization study by: regressing each of the K risk factors against the J genetic variants G_j $(j = 1, \ldots, J)$ to obtain the predicted values of the risk factors ($\hat{X}_1$, $\hat{X}_2$ and $\hat{X}_3$); and then regressing the outcome Y against $\hat{X}_1$, $\hat{X}_2$ and $\hat{X}_3$. The estimates for $\hat{X}_1$, $\hat{X}_2$ and $\hat{X}_3$ from the second stage regression model should be consistent estimates of $\boldsymbol{\theta}$ if the IV assumptions for multivariable Mendelian randomization are satisfied.

Summary level data

We assume that we have summary level data for the three risk factors and outcome, i.e. the genetic associations with the three risk factors are estimated in one sample, and the genetic associations with the outcome are estimated in an independent sample. Consistent estimates of $\boldsymbol{\theta}$ can be obtained from the multivariable weighted linear regression of the genetic association estimates with the K risk factors and the genetic association estimates with the outcome, with $\text{se}(\hat{\beta}_{Y_j})^{-2}$ as weights and the intercept set to zero (known as the 'multivariable IVW method') [80]. Assuming there are three risk factors, under the multivariable IVW method we consider:

$$\hat{\beta}_{Y_j} = \theta_{1MV}\hat{\beta}_{X_{1j}} + \theta_{2MV}\hat{\beta}_{X_{2j}} + \theta_{3MV}\hat{\beta}_{X_{3j}} + \epsilon_{MV_j}, \quad \epsilon_{MV_j} \sim \mathcal{N}(0, \phi^2_{MV}\,\text{se}(\hat{\beta}_{Y_j})^2)\,, \quad (2.13)$$

where $\boldsymbol{\theta_{MV}}$ are the causal effects of the risk factors X_k $(k = 1, \ldots, 3)$ on the outcome Y, when conditioned on each other and ϕ_{MV} represents the residual standard error

under the multivariable IVW model. If $K = 1$, then the multivariable IVW model is equivalent to the 'univariable' IVW method in Equation 2.7.

There may be circumstances where the risk factors $\boldsymbol{X}$ are linearly related. For example, suppose the risk factors under investigation in Figure 2.4 are low-density lipoprotein cholesterol (LDL-C), triglycerides and high-density lipoprotein cholesterol (HDL-C). LDL-C is rarely measured directly, but is estimated from measurements of total cholesterol, triglycerides and HDL-C via the Friedewald equation as total cholesterol minus HDL-C minus 0.2 times triglycerides (assuming all measurements are in mg/dL) [81]. We would therefore expect LDL-C, triglycerides and HDL-C measurements to be correlated, and since lipid fractions are associated with common genetic variants, we may find that the estimates of the genetic associations ($\hat{\boldsymbol{\beta}}_{\boldsymbol{X}}$) are also correlated. If the estimates $\hat{\boldsymbol{\beta}}_{\boldsymbol{X}}$ are correlated, then the multivariable IVW method (Equation 2.13) may be effected by collinearity, leading to imprecise estimates. Collinearity in multivariable Mendelian randomization is considered briefly in the main applied example of the dissertation (Chapter 6) by estimating the correlation structure between the risk factors and the correlation structure between the genetic associations of the risk factors.

Instrument strength

For multivariable Mendelian randomization, the set of genetic variants $\boldsymbol{G}$ are considered to be strong IVs if: a) the variants are associated with all of the K risk factors; and b) the variants are jointly associated with the K risk factors [82]. The first condition can be assessed through the F-statistics from the regression of $\boldsymbol{G}$ against each of the K risk factors X_k ($k = 1, \ldots, K$). In order for the second condition to hold, the genetic variants must be able to predict the values of each risk factor X_k after predicting the values of the remaining $K - 1$ risk factors. In Figure 2.5, all of the risk factors are individually strongly predicted by the J genetic variants G_j ($j = 1, \ldots, J$) for DAG A). X_2 and X_3 are jointly predicted by $\boldsymbol{G}$ in DAG A), but X_1 is not jointly predicted by $\boldsymbol{G}$. In DAG B), all of the risk factors are individually predicted and jointly predicted by $\boldsymbol{G}$.

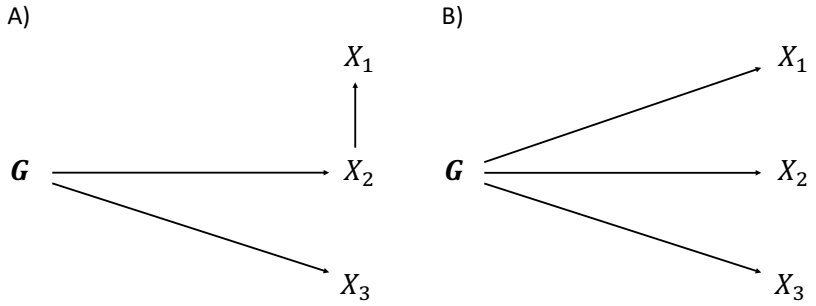


Fig. 2.5 Directed acyclic graph illustrating the potential set up for multivariable Mendelian randomization for the set of genetic variants $\boldsymbol{G}$ and the risk factors X_1, X_2 and X_3. In DAG A), X_1 is individually, but not jointly, strongly predicted by $\boldsymbol{G}$. All of the risk factors in DAG B) are both individually and jointly strongly predicted by $\boldsymbol{G}$.

To assess whether the J genetic variants G_j $(j = 1, \ldots, J)$ are jointly associated with the K risk factors, the Sanderson-Windmeijer conditional F-statistic should be estimated for each risk factor [82]. The conditional F-statistic for X_1 when there are $K = 3$ risk factors can be calculated by:

1. X_2 is regressed against the J genetic variants G_j $(j = 1, \ldots, J)$ and the predicted values of X_2 $(\hat{X}_2)$ calculated;
2. the above regression model is refitted with X_2 replaced with X_3, and $\hat{X}_3$ calculated;
3. X_1 is then regressed against $\hat{X}_2$ and $\hat{X}_3$, and the residual errors from the regression model $(X_1 - \hat{X}_1)$ saved;
4. the saved residual errors are regressed against the J genetic variants G_j $(j = 1, \ldots, J)$, and the F-statistic obtained from this regression model, multiplied by a degrees of freedom correction factor of $J/(J-2)$ [83], is the 'conditional' F-statistic for X_1.

The degrees of freedom correction factor in step 4 takes into consideration that the same set of genetic variants $\boldsymbol{G}$ were used to predict that values of $\hat{X}_2$ and $\hat{X}_3$ in steps 1 and 2. Note that fitting the three regression models in steps 1-3 is equivalent to performing TSLS regression when X_1 is treated as the 'outcome', and X_2 and X_3 are the risk factors. The conditional F-statistics for X_2 and X_3 should also be calculated, and the J genetic variants G_j $(j = 1, \ldots, J)$ would be considered as strong IVs for multivariable Mendelian randomization if the F-statistics and conditional F-statistics for all of the K risk factors X_k $(k = 1, \ldots, K)$ were sufficiently large, e.g. we may apply the traditional cut-off value of 10.

To calculate the Sanderson-Windmeijer conditional F-statistics we require individual level data on the J genetic variants G_j ($j = 1, \ldots, J$) and K risk factors X_k ($k = 1, \ldots, K$). In order to assess the strength of the IVs when only summary level data on the risk factors are available (i.e. in a two–sample Mendelian randomization study), Sanderson *et al.* [82] have proposed an adapted version of the Cochran's Q statistic to test for instrument strength. Although this modified version of the Cochran's Q statistic performed well in the authors' simulation study, the test statistic requires information on the covariance structure between the genetic associations of the K risk factors.

Benefits of multivariable Mendelian randomization

The main benefit of multivariable Mendelian randomization is that it offers an alternative to 'univariable Mendelian randomization' (as considered in Section 2.2), and can be used in the primary or sensitivity analysis. Using multivariable Mendelian randomization may be a good alternative to univariable Mendelian randomization if: a) the IV2 or IV3 assumptions for univariable Mendelian randomization are suspected to be violated; and/or b) the risk factor under consideration is known to be correlated with other risk factors, and the causal effect that all of these risk factors have on the outcome want to be investigated. For either case, assumptions about the relationships between the genetic variants, risk factors and outcome must be made, and this should be informed by biological evidence.

2.7 Conclusion

Ideally, a genetic variant should only be included in a Mendelian randomization analysis if its biological function is well understood. Although this should reduce the risk of including pleiotropic genetic variants, only considering variants with known biological mechanisms would severely limit the scope of Mendelian randomization. To provide robustness to the results from the main analysis, methods that account for pleiotropic genetic variants must be considered in the sensitivity analysis of a Mendelian randomization study.

Section 2.6 highlighted the range of sensitivity analyses that may be used in Mendelian randomization to account for pleiotropic genetic variants using summary level data. The majority of these methods focus on identifying and removing pleiotropic genetic variants from the analysis. Although the multivariable IVW method was developed to account for measured pleiotropy, there are no methods that can be used

as a sensitivity analysis under a multivariable framework. The application of the MR-Egger method to the multivariable setting in Chapter 4 should help to rectify this gap in the literature.

Since the IVW method is equivalent to performing a meta-analysis of the causal ratio estimates, a lot of the sensitivity analyses and tests for heterogeneity among the causal ratio estimates discussed in Sections 2.5 and 2.6 have been adapted from the meta-analysis literature [48, 63, 29]. Since pleiotropic genetic variants may appear as outlying data points in a Mendelian randomization analysis, we suspected that some of the methods in the robust statistics [84] literature that try to reduce the influence outlying data points have on the analysis may be a useful addition to Mendelian randomization (considered in Chapter 3). Some of these robust methods, such as Lasso penalization [65, 67, 85], have already been adapted to Mendelian randomization with individual level data. Since the use of summary level data in Mendelian randomization continues to grow in popularity, the application of methods like Lasso penalization to summary level data should be considered.

Chapter 3

Downweighting or removing heterogeneous causal estimates: robust methods for Mendelian randomization with multiple genetic variants

3.1 Introduction

If the genetic variants in a Mendelian randomization study are all valid IVs, then the individual causal ratio estimates should be similar. Pleiotropic genetic variants are likely to have heterogeneous causal ratio estimates. In the two applied examples considered in this Chapter, we found that heterogeneity of the causal estimates may be considered under two scenarios: 1) when there is over–dispersion in the estimates as there is more variance between the variant specific causal estimates than expected by chance (as seen in the effect of body mass index on schizophrenia); and 2) when specific variants have outlying causal estimates, and they alone are responsible for driving the observed heterogeneity (as seen in the effect of low–density lipoprotein on Alzheimer's disease).

There are numerous methods in the Mendelian randomization literature that detect or account for pleiotropic variants using summary level data. As highlighted in Section 2.6, these methods can be divided into two broad categories: methods that detect pleiotropic genetic variants and either downweight or remove them; and methods

that estimate consistent causal effects in the presence of pleiotropic genetic variants without downweighting their contribution to the causal estimate. In this Chapter, we focus on identifying 'robust methods' for summary level data that either downweight or remove genetic variants with heterogeneous causal ratio estimates. We consult the literature on robust statistics and recent developments in Mendelian randomization to identify such methods. The methods identified in this Chapter will be applied to the two examples highlighted above (the effect of body mass index on schizophrenia, and the effect of low-density lipoprotein on Alzheimer's disease), and we anticipate that these methods will be used in our investigation of the effect of adiposity and body composition on asthma in Chapter 6.

In Section 3.2, we provide a brief overview of the fundamental aims of robust statistics, and discuss why some of the methods in the robust statistics literature may be relevant to Mendelian randomization when there is heterogeneity among the causal ratio estimates. In Section 3.3, we introduce two additional robust methods from the robust statistics literature (robust regression (MM-estimation) and least trimmed squares selection), and outline a selection procedure based on Lasso regression and recent developments in Mendelian randomization [67]. We also outline a robust method that uses the Q-statistic to penalize genetic variants with heterogeneous causal ratio estimates. We apply the methods introduced in Section 3.3 to published data on body mass index and schizophrenia risk, and on low-density lipoprotein and Alzheimer's disease risk (Section 3.4). In Section 3.5, we perform a simulation study to compare bias and coverage properties of the robust methods when some of the genetic variants are invalid IVs. Finally, we discuss the results of the Chapter and its implications to applied Mendelian randomization research (Section 3.6).

All of the computational work for the applied examples (Section 3.4) and the simulation study (Section 3.5) was written and performed by Jessica Rees in RStudio version 3.5.3 [86] unless explicitly stated otherwise. Details on the packages and libraries used in RStudio are provided throughout the Chapter.

3.2 Robust statistics and Mendelian randomization

In this Section, we provide a brief overview of the fundamental aims of robust statistics (Section 3.2.1), discuss the relevance and possible merits of using methods from the robust statistics literature in Mendelian randomization (Section 3.2.2), and outline the robust statistics methods considered in Section 3.3 (Section 3.2.3).

3.2.1 Overview of robust statistics

The primary aim of robust statistics is to develop estimators that are not severely affected by outliers (data points that differ from the majority) or deviations from the model assumptions. Robust statistics may estimate location parameters, scale parameters, and regression coefficients. In this Chapter, we focus on robust statistics methods for estimating regression coefficients where outlying data points may differ from the majority.

Robust statistics may act as replacements to popular statistical estimators that are sensitive to outliers or deviations from the model assumptions. Robust statistics should be efficient, with small deviations in the model assumptions having a minor impact on the performance of the estimator [84]. A balance between the efficiency and stability of a robust statistic must be met. For example, we may prefer a less efficient estimator if it is more likely to be unbiased in the presence of outliers.

A popular measure of robustness of an estimator is the 'breakdown point' [87]. This represents the proportion of 'bad observations', such as outliers, that an estimator can cope with before it produces an arbitrarily large or small result [84]. For example, the ordinary least squares (OLS) estimator has a breakdown point of 0% as one outlying observation can result in a biased OLS estimate, whereas the median has a breakdown point of 50%. Many robust statistics focus on increasing the breakdown point of the estimator by reducing the weight of outliers, including the least trimmed squares estimator and M-estimators considered in Section 3.3.

3.2.2 Relevance of robust statistics to Mendelian randomization

The motivation for considering the robust statistics literature to identify methods for Mendelian randomization was driven by the observation that pleiotropic genetic variants may have outlying (as seen in Figure 2.3 for the effect of low–density lipoprotein on Alzheimer's disease) or over-dispersed (observed for the effect of body mass index on Schizophrenia) summary level data.

As defined in Section 2.4, we assume that the genetic association with the outcome can be expressed as the linear combination of the direct effect (α_j) and the indirect effect via the risk factor ($\theta\beta_{X_j}$):

$$\beta_{Yj} = \alpha_j + \theta\beta_{X_j} \,. \tag{3.1}$$

From Equation 3.1, a pleiotropic genetic variant ($\alpha_j \neq 0$) would be an outlier with respect to the y-axis in the scatter plot of the genetic association estimates with the outcome against the genetic association estimates with the risk factor if the majority of the variants were valid IVs. Since one of the primary objectives of robust statistics is to reduce the impact outliers have on the performance of the estimator, we hypothesised that there may be regression estimators in the robust statistics literature that would produce consistent estimates of the causal effect when pleiotropic genetic variants were included in a Mendelian randomization study.

If there is over–dispersion in the causal ratio estimates, and all of the genetic variants are valid IVs, then the IVW estimate will be consistent, and multiplicative random effects should be used. However, the IVW estimate will be biased if there is over–dispersion and some of the genetic variants are pleiotropic. We anticipated that methods in the robust statistics literature that are less sensitive to data from heavy tailed distributions may be beneficial to Mendelian randomization analyses when there is general over–dispersion in the ratio estimates and there are pleiotropic genetic variants.

3.2.3 Methods considered

There are various estimators in the robust statistics literature that are less sensitive to outliers. Some of these methods perform well in terms of consistency, whilst others focus on enhancing efficiency. We decided to assess the performance of the two robust estimators: MM–estimation and the least trimmed squares estimator (described in detail in Section 3.3). MM–estimation was chosen as it strikes a good balance between efficiency and consistency [88], whereas least trimmed squares is particularly robust to outlying data points, but may lack efficiency [89]. Although our decision to consider these methods may be considered arbitrary, we anticipate that these estimators may provide a good indication as to whether the robust statistics literature is relevant to Mendelian randomization.

3.3 Methods

In this Section, we introduce four robust methods for Mendelian randomization using summary level data: 1) robust regression (MM–estimation); 2) penalized weights; 3) least trimmed squares selection; and 4) Lasso selection. Robust regression (Section 3.3.1) and least trimmed squares selection (Section 3.3.4) are based on the MM-estimator

and least trimmed squares estimator from the robust statistics literature respectively. The penalized weights method uses the Q-statistic to downweight genetic variants with heterogeneous ratio estimates (Section 3.3.2). Finally, the Lasso selection method (Section 3.3.3) is based on Lasso regression and recent developments in Mendelian randomization [67]. These four methods use summary level data of the genetic associations with the risk factor ($\hat{\beta}_{Xj}$, se($\hat{\beta}_{Xj}$)) and the outcome ($\hat{\beta}_{Yj}$, se($\hat{\beta}_{Yj}$)) for the J genetic variants G_j ($j = 1, \ldots, J$).

Throughout the Chapter, we assume linearity and no effect modification of the causal effect of the risk factor on the outcome, and the associations of the genetic variants $\boldsymbol{G}$ (bold variables represent vectors) with the risk factor and with the outcome (as considered in Section 2.2). We also assume that the outcome is a continuous variable, and all of the genetic variants are independent (not in linkage disequilibrium). The application of the methods introduced in this Chapter to binary outcomes is discussed in Section 3.3.5.

3.3.1 Robust regression (MM–estimation)

Before we introduce the robust regression method used in this Chapter, we first consider M– and S–estimators in relation to estimating the regression coefficients $\boldsymbol{\beta}$ in the linear regression model:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \epsilon_i \\ &= \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i \,, \end{aligned}$$

where $i = 1, \ldots, N$, and the error term ϵ_i has an expected value of zero and scale σ^2.

An M–estimator minimises:

$$\sum_{i=1}^{N} \rho\left(\frac{y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}}{\hat{\sigma}}\right), \tag{3.2}$$

where ρ is an objective function and $\hat{\sigma}$ is a scale estimate for the error term. M-estimates for $\boldsymbol{\beta}$ are obtained by taking the partial derivatives of the objective function in Equation 3.2, and solving the system of equations:

$$\sum_{i=1}^{N} \psi\left(\frac{r_i}{\hat{\sigma}}\right) \boldsymbol{x_i} = \sum_{i=1}^{N} \psi(u_i) \boldsymbol{x_i} = \boldsymbol{0} \,, \tag{3.3}$$

where ψ is proportional to the derivative of ρ, and $r_i = y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$. By substituting the weighting function:

$$w(u) = \frac{\psi(u)}{u}, \tag{3.4}$$

into Equation 3.3, and applying an iteratively reweighted least squares (IRLS) algorithm, M-estimates $\hat{\boldsymbol{\beta}}_{\boldsymbol{M}}$ can be obtained. To reduce the impact outliers have on the scale estimate, the median absolute deviation of the residuals can be considered:

$$\hat{\sigma} = \frac{\text{median}(|r_i|)}{0.6745}, \tag{3.5}$$

where median$(|r_i|)$ is multiplied by $1/0.6745$ as the expected value of median$(|r_i|)$ is 0.6745 if the residuals are normally distributed [90]. Note that σ is re-estimated at each iteration until the estimates $\hat{\boldsymbol{\beta}}_{\boldsymbol{M}}$ converge.

If we assume that the error term is independently and normally distributed $\epsilon_i \sim \mathcal{N}(0, 1)$, and Equation 3.2 is set to $\sum_{i=1}^{N} r_i^2$, then the M-estimator is equivalent to the OLS estimator. However, we may want to use an objective function that is less sensitive to outliers, such as Tukey's bisquare objective function:

$$\rho(u_i) = \begin{cases} \frac{c^2}{6}\left(1 - \left[1 - \left(\frac{u_i}{c}\right)^2\right]^3\right) & \text{if } |u_i| < c \\ \frac{c^2}{6} & \text{if } |u_i| \geq c \end{cases}, \tag{3.6}$$

with its weighting function:

$$w(u_i) = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2 & \text{if } |u_i| < c \\ 0 & \text{if } |u_i| \geq c \end{cases}. \tag{3.7}$$

From Equation 3.7, the weight of an observation decreases as u_i tends away from zero, and when $|u_i| \geq c$ the observation will have zero weight. The value of the tuning parameter c determines the relative efficiency of the M–estimator. For Tukey's bisquare weighting function, a standard value of $c = 4.685$ is used to ensure the M-estimator has 95% asymptotic efficiency relative to the OLS estimate if the error term is normally distributed with an expected value of zero and constant variance [89]. Whilst this estimator may be less sensitive to outlying data points with respect to the y observations, it can be sensitive to leverage points (data points that are outlying with respect to $\boldsymbol{x}$). As such, M-estimators may be highly efficient, but lack robustness.

An M–estimate of the scale of the error term is the value $\hat{\sigma}$ that solves:

$$\frac{1}{N}\sum_{i=1}^{N}\rho\left(\frac{y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}}{\sigma}\right) = K\,,$$

where K is a tuning parameter, and ρ is an objective function. S–estimates $\hat{\beta}_S$ are the values that minimise the M–estimate of scale $\hat{\sigma}_S$. Estimates of $\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}$ and $\hat{\sigma}_S$ can be obtained by using the weighting function in Equation 3.4 and an IRLS algorithm. If Tukey's bisquare objective function is used, the S-estimator will have an asymptotic breakdown point of 50% if $c = 1.548$ and $K = 0.5$ [91]. Whilst the S-estimator may be highly robust to outliers and leverage points, it can lack efficiency.

To overcome some of the disadvantages of using the M– and S–estimators, MM–estimation was proposed by Yohai [92], and consists of the following three stages:

1. The initial estimates $\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}$ are obtained from a S-estimator with a high breakdown point and objective function ρ_1.
2. Using the residuals $r_i = y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}$ from the stage above, an M-estimate of scale $\hat{\sigma}_S$ is calculated.
3. M–estimates $\hat{\boldsymbol{\beta}}$ are obtained using the M-estimate of scale $\hat{\sigma}_S$ from the second stage and the objective function ρ_2. Note that the M-estimate of scale $\hat{\sigma}_S$ is fixed for each iteration, i.e. it is not re-estimated using Equation 3.5. The estimates $\hat{\boldsymbol{\beta}}$ from this stage represent the MM-estimates.

By using a S-estimator in the first stage, and M-estimator in the third stage, the MM-estimator should be efficient and have a high breakdown point.

In this Chapter, we consider the MM–estimation approach (referred to as 'robust regression' in this dissertation) described by Yohai [92] and Koller and Stahel [88] that is used by the `lmrob` command in the R package *robustbase* [93]. `lmrob` uses Tukey's bisquare objective function (Equation 3.6) for ρ_1 and ρ_2, with $c = 1.548$ in Equation 3.7 at the S–estimation step to maintain a high breakdown point, and $c = 4.685$ at the M–estimation step to provide efficiency.

Instead of using weighted least squares to obtain the estimates for the IVW and MR-Egger methods (as done in Equations 2.7 and 2.9), we propose using robust regression (MM–estimation approach used by `lmrob`). Since the `lmrob` command allows the user to specify a vector of weights to be used in conjunction with Tukey's weighting function, we also account for $\text{se}(\hat{\beta}_{Y_j})^{-2}$ in the weights of the observations. The estimates from the MM-estimator using Tukey's weighting function $w(u_j)$ and $\text{se}(\hat{\beta}_{Y_j})^{-2}$ as weights, is

equivalent to the estimates obtained from weighted least squares where the weights are the product of $w(u_j)$ from the final iteration of the MM-estimator and $\text{se}(\hat{\beta}_{Y_j})^{-2}$.

3.3.2 Penalized weights

As highlighted in Section 2.5.2, Cochran's Q-statistic has been adapted from the meta-analysis literature to test for heterogeneity among the ratio estimates for the IVW method:

$$Q = \sum_j Q_j = \sum_j w_j(\hat{\theta}_j - \hat{\theta})^2 \,, \tag{3.8}$$

where Q has an approximate χ^2_{J-1} distribution under the null hypothesis that all J genetic variants satisfy the IV assumptions, with the J components Q_j $(j = 1, \ldots, J)$ having approximate χ^2 distributions with one degree of freedom [62]. The Q-statistic has also been used in Mendelian randomization to downweight [52] or exclude genetic variants with heterogeneous ratio estimates [63]. The Q-statistic is based on the first order weights (Equation 2.6) of the IVW method which assumes that there is no measurement error (NOME) in the genetic associations with the risk factor [63]. When the Q-statistic is used as a test for heterogeneity, the ratio estimates are usually compared against the IVW estimate ($\hat{\theta} = \hat{\theta}_{IVW}$ in Equation 3.8). If some of the J genetic variants G_j $(j = 1, \ldots, J)$ are pleiotropic, then the IVW estimate will be biased.

By substituting $\hat{\theta}_j = \hat{\beta}_{Yj}/\hat{\beta}_{Xj}$ into Equation 3.8, and using the first order approximation of w_j, we obtain:

$$Q = \sum_j Q_j = \sum_j \text{se}(\hat{\beta}_{Yj})^{-2}(\hat{\beta}_{Yj} - \hat{\theta}\hat{\beta}_{Xj})^2 \,. \tag{3.9}$$

We will use the J components Q_j $(j = 1, \ldots, J)$ of the Q-statistic in Equation 3.9 to downweight genetic variants with heterogeneous ratio estimates. To provide additional robustness, the simple (unweighted) median estimator for $\hat{\theta}$ will be used in Equation 3.9 as it produces consistent causal estimates if at least 50% of the genetic variants are valid IVs [52]. We propose fitting the weighted linear regression of the genetic association estimates with the risk factor $(\hat{\beta}_{X_j})$ and the genetic association estimates with the outcome $(\hat{\beta}_{Y_j})$, with the intercept fixed at zero and weights:

$$\text{se}(\hat{\beta}_{Yj})^{-2} \times \min(1, 100q_j) \tag{3.10}$$

where q_j is the probability of observing a value greater than or equal to Q_j (Equation 3.9) from the χ^2 distribution with one degree of freedom. If $\min(1, 100q_j) = 1$ for all of

the J genetic variants, then all of the weights will remain the same, and the weighted linear regression model will be equivalent to the IVW method (Equation 2.7).

We initially considered a downweighting factor of $\min(1, 20q_j)$ in Equation 3.10 as this was used in the paper by Bowden *et al.* [52] for the penalized–median estimator. We found that too many variants were being penalized, resulting in over-precise estimates that had poor coverage of the true causal effect. By multiplying the first order approximation of the weights by $\min(1, 100q_j)$, the outlying variants that are suspected to be pleiotropic should be severely penalized, without downweighting too many genetic variants.

We will also consider penalizing the weights for the MR-Egger method by using the modified Q' statistic outlined by Bowden *et al.* [94]:

$$Q' = \sum_j Q'_j = \sum_j \text{se}(\hat{\beta}_{Yj})^{-2}(\hat{\beta}_{Yj} - \hat{\theta}_0 - \hat{\theta}_1\hat{\beta}_{Xj})^2\,, \tag{3.11}$$

where $\hat{\theta}_0$ and $\hat{\theta}_1$ are the MR-Egger estimates from the weighted linear regression of the genetic association estimates with the risk factor ($\hat{\beta}_{X_j}$) and the genetic association estimates with the outcome ($\hat{\beta}_{Y_j}$) [45], with the inverse-variance as weights ($\text{se}(\hat{\beta}_{Y_j})^{-2}$) and the intercept unrestrained. If the MR-Egger model is correct, the Q' statistic in Equation 3.11 should follow an approximate χ^2_{J-2} distribution, with the J components Q'_j ($j = 1, \ldots, J$) having approximate χ^2 distributions with one degree of freedom [95]. We propose re-fitting the MR-Egger regression model with the penalized weights:

$$\text{se}(\hat{\beta}_{Yj})^{-2} \times \min(1, 100q'_j) \tag{3.12}$$

where q'_j is the probability of observing a value greater than or equal to Q'_j (Equation 3.11) from the χ^2 distribution with one degree of freedom. If $\min(1, 100q'_j) = 1$ for all of the J genetic variants, then all of the weights will be unchanged, and the MR-Egger estimates will be the same as $\hat{\theta}_0$ and $\hat{\theta}_1$ in Equation 3.11.

The penalized weights in Equations 3.10 and 3.12 will be applied to the IVW and MR-Egger methods with weighted least squares regression. These penalized weights can also be applied to the IVW and MR-Egger method using robust regression (Section 3.3.1). There may be additional robustness against pleiotropic genetic variants by combining robust regression with the penalized weights described above.

3.3.3 Lasso selection

Lasso regression is a regularization method that performs variable selection by estimating regression coefficients subject to a penalty term [96]. This regularization should help to reduce over-fitting, creating a more parsimonious model. Using the notation from Section 3.3.1, the Lasso estimates $\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}$ are the values that minimise:

$$\sum_{i=1}^{N} \rho(x_i, y_i, \beta_0, \boldsymbol{\beta}) + \lambda \sum_{p=1}^{P} |\beta_p| \,,$$

where ρ is the objective function, and $\lambda \sum |\beta_p|$ is the Lasso penalty term (also referred to as the L1 penalty term). The value of the tuning parameter λ will determine the amount of shrinkage applied to the Lasso estimates.

The Lasso estimator has been considered in the IV literature using indiviudal level data [65, 67, 85]. Windmeijer *et al.* [67] provide an overview of some of the methods that use Lasso in IV analyses. In particular, they describe the 'post-lasso' estimator that uses Lasso to select the genetic variants that will be used as IVs in the main analysis. The post-lasso estimator has been considered in the IV literature prior to the paper by Windmeijer *et al.* [67], for example, by Belloni *et al.* [97]. For individual level data, the post-lasso estimator consists of two stages: 1) the Lasso estimator selects the genetic variants that will be used as IVs; and 2) the genetic variants selected in the first stage are used in the TSLS regression model to obtain an estimate of the causal effect.

Since the majority of Mendelian randomization studies use summary level data, we consider the possibility of applying Lasso regression to summary level data to select genetic variants that are used as IVs in the IVW method. We first consider the objective function for the MR-Egger model:

$$\sum_{j} \mathrm{se}(\hat{\beta}_{Yj})^{-2} (\hat{\beta}_{Yj} - \theta_0 - \theta_1 \hat{\beta}_{Xj})^2 \,.$$

We propose replacing θ_0 with a separate intercept coefficient for each genetic variant θ_{0j}, and adding a Lasso penalty term for θ_{0j}:

$$\sum_{j} \mathrm{se}(\hat{\beta}_{Yj})^{-2} (\hat{\beta}_{Yj} - \theta_{0j} - \theta_1 \hat{\beta}_{Xj})^2 + \lambda \sum_{j} |\theta_{0j}| \,. \tag{3.13}$$

A genetic variant is selected as a valid IV if the estimate for θ_{0j} shrinks to zero in Equation 3.13. The genetic variants that are selected as valid IVs (i.e. they have a zero intercept $\hat{\theta}_{0j}$) are included in the IVW model to estimate the causal effect. The

number of genetic variants that are selected for the IVW model is determined by the value of the tuning parameter λ in Equation 3.13. If $\lambda = \infty$, the intercept terms are forced to be zero for all J variants, and the IVW model is fitted to the full set of genetic variants. If $\lambda = 0$, then all of the variants can be pleiotropic, and the IVW model cannot be fitted as none of the genetic variants are selected as valid IVs.

To determine the value of λ, two rules were considered: a heterogeneity stopping rule, and a cross-validation rule. The heterogeneity stopping rule is based on the estimate of the residual standard error $\hat{\phi}$ from the IVW model (Equation 2.7). For the heterogeneity stopping rule, we fit the Lasso regression model (Equation 3.13) over a range of values for λ, starting with a value close to zero, and then increasing λ in small increments. We stop at $\lambda = \lambda_n$ when:

- $\hat{\phi}_{\lambda_{n+1}} > 1$ from the IVW model that used the selected genetic variants from Equation 3.13 when $\lambda = \lambda_{n+1}$; and
- The following condition is satisfied:

$$(\hat{\phi}_{\lambda_{n+1}} - \hat{\phi}_{\lambda_n}) > \frac{\chi^2(0.95)}{J_{inc}},$$

  where $\chi^2(0.95)$ is the upper 95$^{\text{th}}$ percentile of a χ^2 distribution with one degree of freedom, and J_{inc} is the number of genetic variants selected for the IVW model from Equation 3.13 when $\lambda = \lambda_{n+1}$.

The estimate from the IVW model that included the genetic variants selected from Equation 3.13 when $\lambda = \lambda_n$ is treated as the causal estimate. An algorithm for the stopping rule was written by Stephen Burgess in R to determine the value of λ under the heterogeneity stopping rule using the `penalized` command in the R package *penalized*. This R code was used by Jessica Rees in the applied examples (Section 3.4) and simulation study (Section 3.5).

As an alternative to the heterogeneity stopping rule, we use the `optL1` command in the R package *penalized* [98]. `optL1` compares the predictive ability of the Lasso regression model for different values of λ through leave-one-out cross-validation. The optimal value of λ is then determined by maximizing the cross-validated likelihood function. We may find that the values of λ are substantially different for the heterogeneity stopping rule and cross-validation as the stopping rule treats λ as a discrete variable, whereas the cross-validation method takes a more informative approach by treating λ as a continuous variable.

3.3.4 Least trimmed squares selection

OLS regression minimises the sum of the squared residuals over N data points; whereas least trimmed squares (LTS) minimises the sum of the squared residuals over a subset h of the N observations [99]. The LTS estimator minimises the following objective function:

$$\sum_{i=1}^{h}(r^2)_{i:N}\,,$$

where $(r^2)_{1:N}, (r^2)_{2:N}, \ldots, (r^2)_{N:N}$ are the ordered squared residuals. The value of h will determine the breakdown point of the LTS estimator and must satisfy $N/2 < h \leq N$. The LTS estimator will achieve an asymptotic breakdown point of 50% when $h = ([N/2] + [(p+1)]/2)$, where p is the number of covariates included in the model [99]. As the value of h increases, the breakdown point will decrease until it reaches zero at $h = N$, where the LTS regression estimate $\hat{\theta}_{LTS,h}$ will be equivalent to the OLS estimate.

The application of LTS regression has been limited due to its low efficiency and computational time. Methods have been proposed to overcome these issues, in particular, the re-weighted version of the LTS estimator was developed to improve efficiency [100]. Under this method, the LTS scale estimate $\hat{\sigma}$ (the objective function multiplied by a consistency factor) is used to compute robust standardized residuals $r_i/\hat{\sigma}$ for all N observations [101], where the residuals r_i are taken from the LTS regression model based on the h data points. A weighting function is then applied to the standardized residuals where observations are given a weight of zero if $|r_i/\hat{\sigma}| > 2.5$, and 1 otherwise. These weights ($w_{LTS,1}$) are then used to update the scale estimate and obtain a new set of robust standardized residuals. The updated standardized residuals are then applied to the same weighting function to obtain a new set of weights ($w_{LTS,2}$) for the N observations. The second set of weights $w_{LTS,2}$ can then be used in a weighted least squares regression model.

The `ltsReg` command in the R package *robustbase* [93] performs the re-weighted version of LTS regression, with a default value of $([N/2] + [(p+1)]/2)$ for h. The function reports an estimate of the scale parameter and the coefficients from the weighted least squares regression model with $w_{LTS,1}$ as weights. The `ltsReg` function also returns the second set of weights $w_{LTS,2}$ and the h observations used to obtain the initial LTS estimate $\hat{\theta}_{LTS,h}$.

We propose using the LTS estimator to select the genetic variants used in the IVW model. The estimate from the IVW model will be consistent if at least 50% of the genetic variants are valid IVs, and the LTS estimator correctly identifies the valid

IVs. The following three sets of variants will be selected using the LTS estimator and will used as IVs in the IVW model: 1) the h variants (approximately 50% of the data) used to estimate the initial LTS estimate $\hat{\theta}_{LTS,h}$; 2) the variants with a weight of 1 in $w_{LTS,2}$; and 3) the variants selected from an automated approach. Under the automated method, the value of h is increased to include more variants in the LTS regression model. h is first set to $h_1 = (J/2 + 1)$ and increased in increments of 1 up to J. For each value of h, the IVW model is fitted to the h variants selected by the LTS objective function, and the residual standard error from the IVW model is recorded. We apply the heterogeneity stopping rule outlined for Lasso selection in Section 3.3.3 to the residual standard errors to determine the optimal value of h.

3.3.5 Binary outcomes

The robust methods in this Section have been introduced under the assumption of linearity and no effect modification in the causal effect of the continuous risk factor X on the continuous outcome Y, and the associations of the genetic variants $\boldsymbol{G}$ with X and Y. Hence, the genetic associations with the risk factor β_{X_j} $(j = 1, \ldots, J)$ or the outcome β_{Y_j} $(j = 1, \ldots, J)$ can be estimated by regressing the risk factor or outcome against each of the genetic variants in linear regression models, where it is assumed that the minor allele has an additive effect on the risk factor or outcome.

We note that the applied examples in Section 3.4, and the main applied example of the dissertation (Chapter 6), all consider binary outcomes. When Y is binary, the genetic association estimates $\hat{\beta}_{Y_j}$ $(j = 1, \ldots, J)$ represent the log odds ratios from a logistic regression model. As discussed in Section 2.3.3, the IVW method will provide an approximate measure of the causal odds ratio for the effect of a continuous risk factor X on a binary outcome Y. The performance of the robust methods introduced in this Section when the outcome is binary will be considered in the simulation study (Section 3.5). Note that the estimates presented in the applied examples in Section 3.4 represent estimates of the approximate log causal odds ratios.

3.3.6 Summary and overview of methods

In this Section, we have introduced robust methods that can be used in a Mendelian randomization study as part of the sensitivity analysis. Table 3.1 provides an overview of the methods, and indicates whether these methods can be applied to the IVW method and/or the MR-Egger method. The methods have been categorised according to whether they downweight genetic variants with heterogeneous ratio estimates whilst

Table 3.1 Overview of the robust methods introduced in Chapter 3. The methods are categorised by whether they downweight genetic variants with heterogeneous ratio estimates (can be applied to the IVW or MR-Egger methods) or select genetic variants for the IVW method.

	IVW method	MR-Egger method
Downweighting heterogeneous ratio estimates		
Robust regression (Rr)	✓	✓
Penalized weights (PW)	✓	✓
Robust regression and penalized weights (Rr and PW)	✓	✓
Selecting instrumental variables		
Lasso selection (LS)		
Heterogeneity stopping rule	✓	-
Cross validation	✓	-
Least trimmed squares (LTS) selection		
h variants	✓	-
$w_{LTS,2}$ variants	✓	-
Automated approach (Auto)	✓	-

keeping them in the analysis model (robust regression and penalized weights), or the method selects the genetic variants used as IVs in the IVW method (Lasso and LTS selection). Table 3.1 contains the abbreviations used in the result tables for the simulation study.

Robust regression and penalized weights can be used separately or in combination with each other, and may be applied to the IVW and MR-Egger methods (Table 3.1). The Lasso selection method uses the heterogeneity stopping rule or cross validation to select genetic variants for the IVW method, and LTS selection identifies three sets of variants for the IVW method.

In the next Section, we apply the robust methods in Table 3.1 with respect to the IVW method only using published summary level data to investigate the two applied examples outlined in Section 3.1: the effect of body mass index on schizophrenia risk, and the effect of low-density lipoprotein on Alzheimer's disease risk.

3.4 Applied examples

To illustrate the performance of the methods proposed in Section 3.3 (summarised in Table 3.1), we considered two applied examples: one where there was evidence of over-dispersion in the ratio estimates (the effect of body mass index (BMI) on schizophrenia risk); and another that contained outliers (the effect of low-density lipoprotein (LDL-C) on Alzheimer's disease (AD) risk). Using summary data (beta–coefficients and standard errors) from PhenoScanner [26], we considered the IVW method with: 1) the full set of genetic variants; 2) robust regression; 3) penalized weights; 4) robust regression and penalized weights; 5) the two sets of genetic variants from Lasso selection using the heterogeneity stopping rule and cross-validation; and 6) the three sets of genetic variants from LTS selection as outlined in Section 3.3.4. Under the heterogeneity stopping rule, the Lasso selection model was applied to $\lambda = 0.1, 0.2, \ldots, 4.9, 5.0, 5.2, 5.4, \ldots, 9.8, 10.0$. Hence, the robust methods in Table 3.1 with respect to the IVW method were applied to the two examples. For comparison with the methods introduced in this Chapter, estimates from the simple median, weighted median, and MR-Egger methods were obtained. Multiplicative random-effects models were used in all analyses.

The genetic association estimates from PhenoScanner for the two binary outcomes (schizophrenia and AD) are the log odds ratios, whereas the genetic association estimates for the risk factors (BMI and LDL-C) are from linear regression models. As highlighted in Section 3.3.5, the results in this Section are estimates of the approximate log causal odds ratios. The simulation study (Section 3.5) will consider how well the robust methods approximate the causal odds ratio.

3.4.1 Causal effect of body mass index on schizophrenia risk

Although individuals with schizophrenia tend to be over-weight [102], it is generally believed that this is due to the effect of anti-psychotic medication on body composition (reverse causation) rather than any causal effect of BMI on schizophrenia risk [103]. For the Mendelian randomization analysis, we used the 97 genetic variants reported by the Genetic Investigation of Anthropometric Traits (GIANT) consortium that were associated with BMI in 339,224 European-descent individuals at a genome-wide level of significance (p-value $< 5 \times 10^{-8}$) [104]. The genetic associations with schizophrenia were obtained from the Psychiatric Genomics Consortium (PGC) based on 35,476 cases and 46,839 controls mostly of European descent [105]. The summarized data used in this Chapter was recently applied to a Mendelian randomization study investigating the causal effect of BMI on psychiatric disorders, including schizophrenia risk [106].

3.4.2 Causal effect of low-density lipoprotein on Alzheimer's disease risk

Epidemiological studies have provided evidence of an association between LDL-C and increased risk of AD [107, 108]. However, there is also evidence to suggest that patients with AD have altered lipid metabolism (reverse causation) [109]. In the Mendelian randomization analysis, we used the 75 genetic variants previously demonstrated to be associated with LDL-C at a genome-wide level of significance by the Global Lipids Genetics Consortium (GLGC) [110]. The point estimates for the genetic associations with LDL-C were taken from the linear regression in up to 188,578 participants from GLGC [111]. A recent Mendelian randomization study used summarized data from GLGC to investigate the causal association between low LDL-C levels and AD risk using data on 380 variants. Our analysis is based on a smaller set of genetic variants compared to Benn *et al.* [112] as we excluded variants that were associated with LDL-C and high density lipoprotein and/or triglycerides. The genetic associations with AD were obtained from the International Genomics of Alzheimer's Project (IGAP) based on 17,008 cases and 37,154 controls of European-descent [113].

3.4.3 Results

The estimated genetic associations with 95% confidence intervals for the two examples are displayed in Figure 3.1. The plots highlight the over-dispersion in the ratio estimates for BMI and schizophrenia; and two outliers in the LDL-C and AD example. The outlying variants (rs6859 and rs7254892) for LDL-C and AD are located near to the *APOE* locus and are associated with AD risk with odds ratios of 1.40 (95% CI: 1.35, 1.44) and 1.28 (95% CI: 1.15, 1.44) respectively [113].

Estimates and 95% confidence intervals from the Mendelian randomization analyses are provided in Table 3.2. All of the estimates for BMI and schizophrenia suggest a null causal effect (as also observed in the Mendelian randomization study by Hartwig *et al.* [106]), although there is wide variation in the standard errors. The use of penalized weights and robust regression in the IVW method improved the precision of the estimates. There was little difference in the point estimates or standard errors obtained from the IVW method with penalized weights, and from the IVW method with robust regression and penalized weights. With exception of the IVW and MR-Egger methods, the unweighted median estimate was the least precise, and including weights in the median estimator had little impact on the standard error of the estimate.

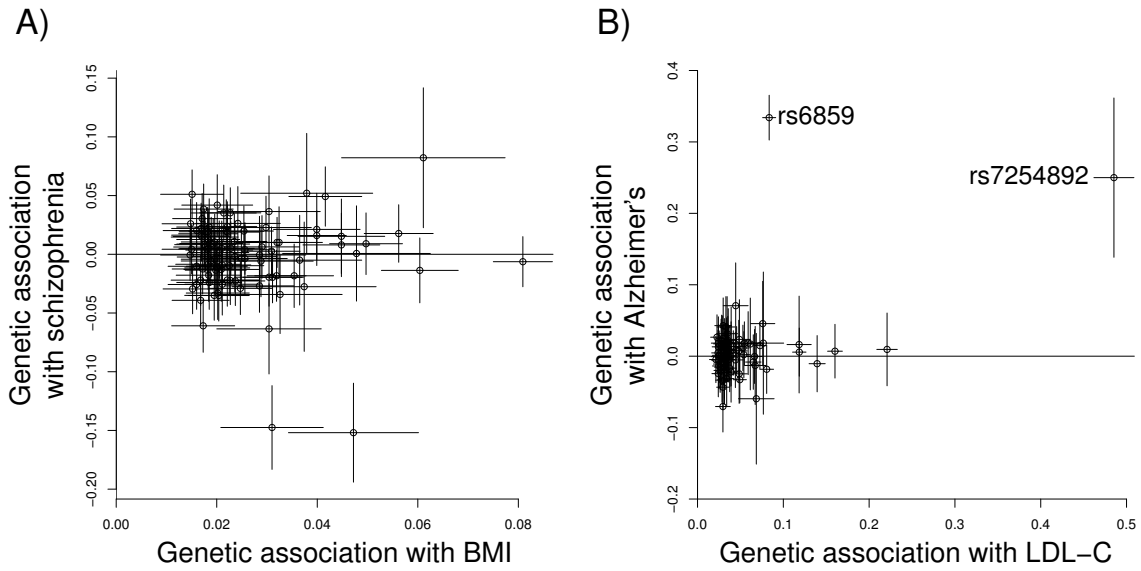


Fig. 3.1 Graph A) contains the estimated genetic associations and 95% confidence intervals with body mass index (BMI, standard deviation units) and with schizophrenia (log odds ratios) for 97 genetic variants. Graph B) contains the estimated genetic associations and 95% confidence intervals with low-density lipoprotein (LDL-C, standard deviation units) and with Alzheimer's disease (log odds ratios) for 75 genetic variants: the two outlying variants are labelled with their rsID codes.

The Lasso selection estimates are displayed in Figure 3.2 (R code for the diagram written by Stephen Burgess) where the approximate causal estimates are relatively similar across the different values of the tuning parameter. The value of the tuning parameter λ was 1.9 under the heterogeneity stopping rule, with 64 genetic variants included in the IVW model. The cross-validation method returned a much larger value of $\lambda = 6.63$, with 95 of the 97 variants included in the IVW model.

Despite using different sets of genetic variants, all of the estimates from LTS selection suggested a null causal effect between BMI and schizophrenia risk. 49 variants were included in the IVW model when LTS selection was based on the h variants used to obtain the initial LTS estimate. 92 variants were included in the IVW model when $w_{LTS,2}$ was used to select the genetic variants, and under the automated approach, 90 variants were included. The point estimate obtained from the h variants was noticeably more precise than the estimates obtained from the other two sets of variants.

The estimates from the IVW, LTS selection with h variants (38 genetic variants included in the IVW model), and MR-Egger methods suggested a positive causal effect of LDL-C on AD risk. This effect was attenuated to the null for the other robust methods. Compared to the robust methods that reported a null causal effect of LDL-C on AD risk, the simple and weighted median estimates had larger standard errors. The

estimates from the IVW and MR-Egger methods from Benn *et al.* [112] indicated that lower LDL-C levels may be beneficial in reducing AD risk, whereas their estimate from the weighted median method suggested a null effect. Since the genetic variants in the *APOE* gene region tend to be highly pleiotropic [73], it is likely that the positive effects obtained from the IVW models in our analysis and in the paper by Benn *et al.* [112] are driven by these pleiotropic variants, rather than there being a true causal effect of LDL-C on AD risk.

None of the estimates in Figure 3.3 (R code for the diagram written by Stephen Burgess) include information on the rs6859 variant, and this outlying variant was only included in the IVW model for Lasso selection when $\lambda = 19.8$, whereas the other outlying variant (rs57254892) was included when $\lambda = 3.5$. Even though the rs57254892 variant seems to be an obvious outlier in Figure 3.1, the genetic variant was included in the IVW model under Lasso selection before many of the other genetic variants. This observation highlights the importance of considering robust methods to identify outlying variants rather than solely relying on plots such as Figure 3.1 to detect outliers 'by eye'.

The λ values for the heterogeneity stopping rule ($\lambda = 3.4$ based on 72 genetic variants) and cross-validation ($\lambda = 4.00$ based on 73 genetic variants) for Lasso selection were similar (Figure 3.3). However, the estimate based on 72 genetic variants was much closer to the null, demonstrating the sensitivity of the IVW method to a single variant. 70 variants were included in the IVW model when $w_{LTS,2}$ was used to select the genetic variants under LTS selection, and 72 variants were used under the automated approach. These sets of variants produced very similar estimates to Lasso selection with the heterogeneity stopping rule.

3.4.4 Summary

In this Section, we applied the methods outlined in Section 3.3 to published summary level data to assess the causal effect of BMI on schizophrenia risk, and the causal effect of LDL-C on AD risk. The consistency of the results from the robust methods for the BMI and schizophrenia example strengthened the evidence from the primary IVW analysis. The LDL-C and AD example highlighted the possibility that only using the IVW method may provide conclusions that are not representative of the majority of the data. Whilst in practice the outlying rs6859 variant could have been identified and removed from the dataset prior to the analysis, the robust methods identified this outlying variant in an automated manner. In the next Section, we perform a simulation

study to assess the performance of the robust methods, and compare the results to the IVW, median, and MR-Egger methods.

Table 3.2 Estimates (standard errors) and 95% confidence intervals of the approximate causal effect of body mass index on schizophrenia risk (log odds ratio for schizophrenia per 1 standard deviation increase in body mass index) and low-density lipoprotein on Alzheimer's disease risk (log odds ratio for Alzheimer's per 1 standard deviation increase in low-density lipoprotein) from the IVW method with: 1) the full set of genetic variants (IVW); 2) robust regression; 3) penalized weights; 4) robust regression and penalized weights; 5) the two sets of genetic variants from Lasso selection using the heterogeneity stopping rule and cross-validation; and 6) the three sets of variants selected under the LTS selection method. Results from the simple median, weighted median and MR-Egger methods are also presented.

	Estimate (SE)	95% CI
Applied example 1: Causal effect of BMI on schizophrenia risk		
IVW	-0.031 (0.100)	-0.227, 0.165
Robust regression	-0.024 (0.079)	-0.180, 0.132
Penalized weights	-0.056 (0.065)	-0.184, 0.073
Robust regression with penalized weights	-0.052 (0.066)	-0.182, 0.078
Lasso selection		
Heterogeneity stopping rule	-0.022 (0.055)	-0.131, 0.086
Cross validation	-0.036 (0.087)	-0.207, 0.136
LTS selection		
h variants	0.077 (0.060)	-0.042, 0.195
$w_{LTS,2}$ variants	-0.042 (0.083)	-0.205, 0.121
Automated approach	-0.056 (0.079)	-0.211, 0.099
Median		
Simple	-0.073 (0.083)	-0.237, 0.090
Weighted	-0.075 (0.090)	-0.252, 0.102
MR-Egger	0.336 (0.241)	-0.136, 0.808
Applied example 2: Causal effect of LDL-C on AD risk		
IVW	0.239 (0.102)	0.039, 0.439
Robust regression	0.048 (0.038)	-0.027, 0.123
Penalized weights	0.040 (0.042)	-0.043, 0.123
Robust regression with penalized weights	0.046 (0.032)	-0.016, 0.108
Lasso selection		
Heterogeneity stopping rule	0.032 (0.044)	-0.054, 0.118
Cross validation	0.088 (0.045)	0.000, 0.175
LTS selection		
h variants	0.172 (0.072)	0.032, 0.313
$w_{LTS,2}$ variants	0.029 (0.043)	-0.054, 0.113
Automated approach	0.032 (0.044)	-0.054, 0.118
Median		
Simple	0.108 (0.071)	-0.031, 0.247
Weighted	0.046 (0.061)	-0.073, 0.165
MR-Egger	0.391 (0.168)	0.061, 0.722

Abbreviations: SE, standard error; CI, confidence interval; BMI, body mass index; IVW, inverse-variance weighted; LTS, least trimmed squares; LDL-C, low-density lipoprotein cholesterol; AD, Alzheimer's disease.

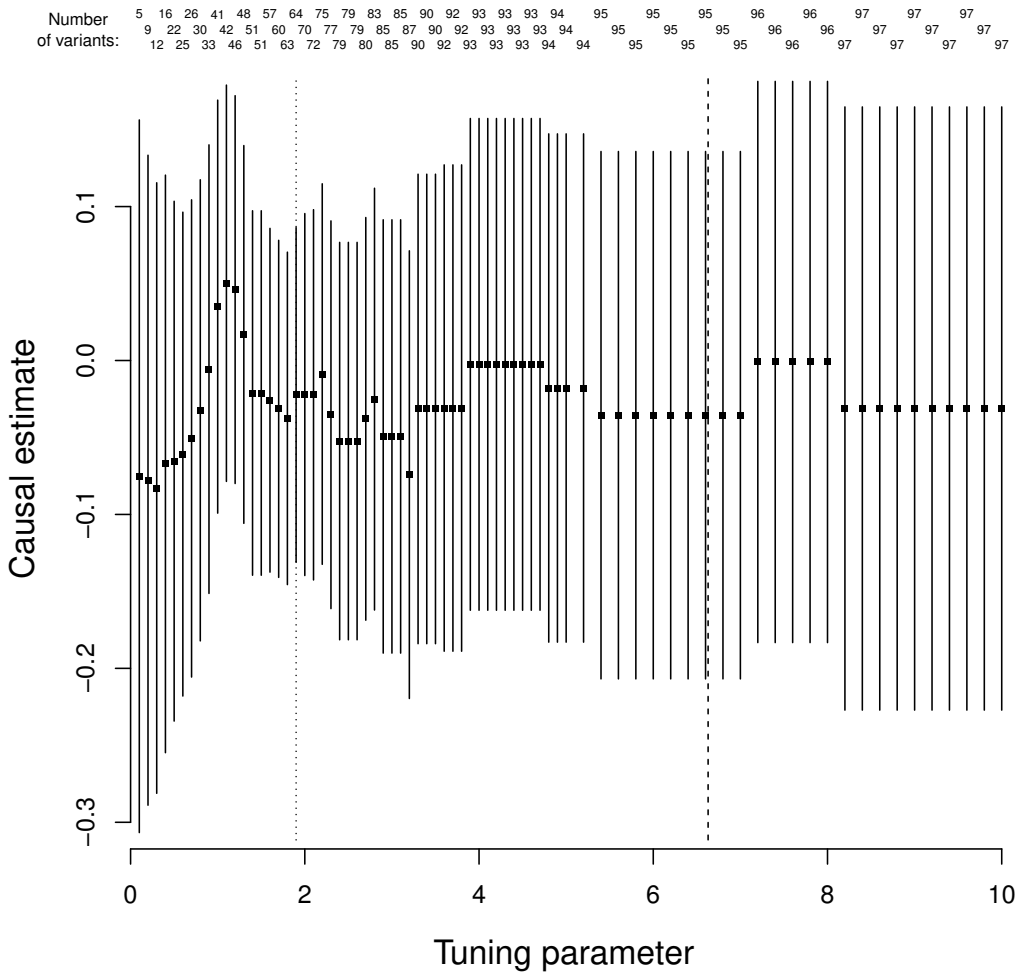


Fig. 3.2 Estimates of the approximate log odds ratio and 95% confidence intervals for schizophrenia per 1 standard deviation increase in body mass index for different values of the tuning parameter ($\lambda = 0.1, 0.2, \ldots, 4.9, 5.0, 5.2, 5.4, \ldots, 9.8, 10.0$) included in the Lasso regression model. The number of genetic variants included in the IVW models are also displayed. The dotted line at $\lambda = 1.9$ is the value of the tuning parameter chosen by the heterogeneity stopping rule. The dashed line at $\lambda = 6.63$ is the value chosen by cross-validation.

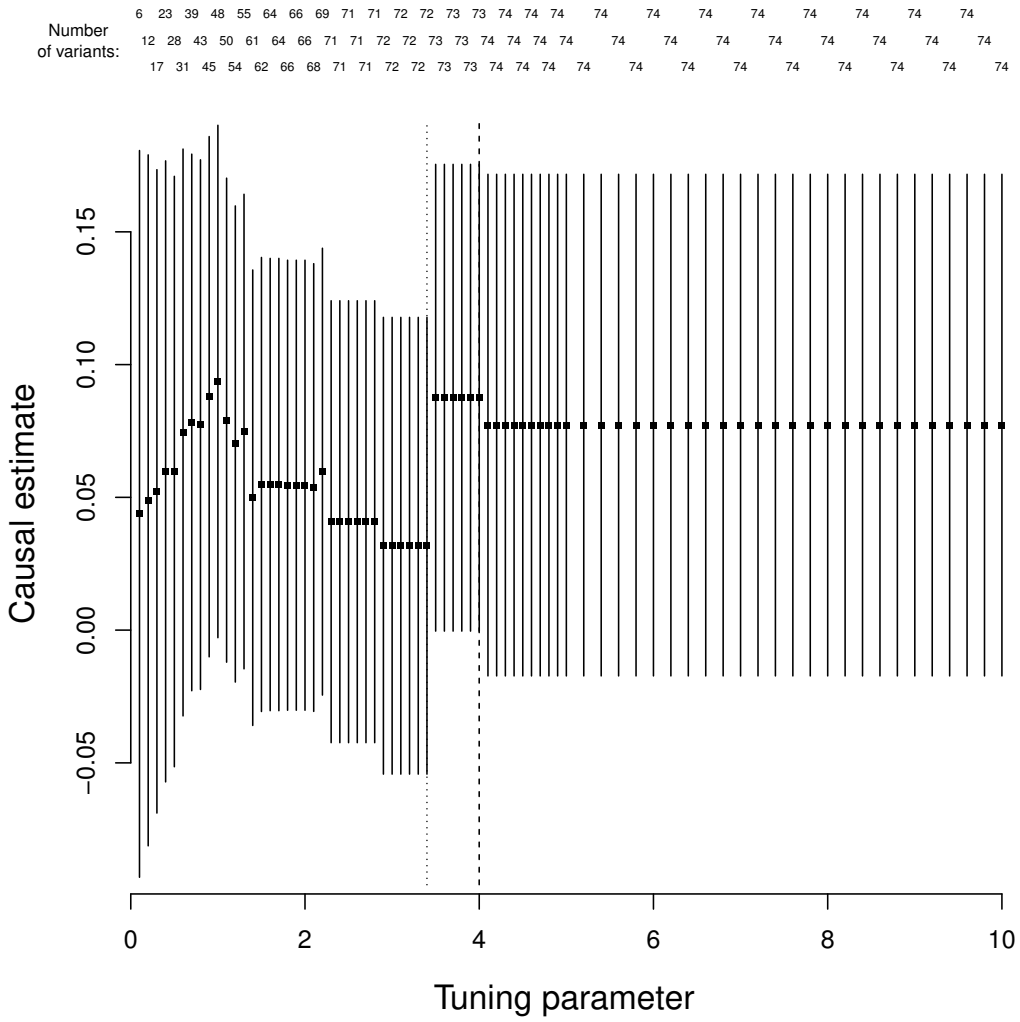


Fig. 3.3 Estimates of the approximate log odds ratio and 95% confidence intervals for Alzheimer's per 1 standard deviation increase in low-density lipoprotein for different values of the tuning parameter ($\lambda = 0.1, 0.2, \ldots, 4.9, 5.0, 5.2, 5.4, \ldots, 9.8, 10.0$) included in the Lasso regression model. The number of genetic variants included in the IVW models are also displayed. The dotted line at $\lambda = 3.4$ is the value of the tuning parameter chosen by the heterogeneity stopping rule. The dashed line at $\lambda = 4.00$ is the value chosen by cross-validation.

3.5 Simulation study

In this Section, we perform a simulation study to compare the performances of the methods outlined in Section 3.3 with the IVW, simple (unweighted) median, weighted median, and MR-Egger methods. To allow for direct comparisons with the MR-Egger method, and to assess the performance of the methods when the IV assumptions are violated, the simulations follow a similar structure to the simulation study performed in the paper by Bowden *et al.* [52]. The data generating model and the methods applied to the simulated data are outlined below.

3.5.1 Data generating model

The simulation study generated data in accordance to Figure 3.4 for participants indexed by $i = 1, \ldots, N$, and genetic variants indexed by $j = 1, \ldots, J$:

$$\begin{aligned}
U_i &= \sum_{j=1}^{J} \delta_j G_{ij} + \epsilon_{Ui}\,, \\
X_i &= \sum_{j=1}^{J} \beta_{X_j} G_{ij} + U_i + \epsilon_{Xi}\,, \\
Y_i &= \sum_{j=1}^{J} \alpha_j G_{ij} + \theta X_i + U_i + \epsilon_{Yi}\,, \\
G_{ij} &\sim \text{Binomial}(2, 0.3) \text{ independently for all } j = 1, \ldots, J\,, \\
\epsilon_{Ui}, \epsilon_{Xi}, \epsilon_{Yi} &\sim \mathcal{N}(0, 1) \text{ independently}\,,
\end{aligned}$$

where α_j represents the direct effect of the genetic variant G_j on the outcome, δ_j represents the effect of the genetic variant on the unmeasured confounder U of the risk factor X and outcome Y association, β_{X_j} represents the genetic effect of G_j on X, and θ is the causal effect of X on Y. The error terms ϵ_{Ui}, ϵ_{Xi}, and ϵ_{Yi} were drawn independently from standard normal distributions.

The performance of the robust methods were investigated under a two-sample Mendelian randomization setting with $N = 10,000$ individuals and $J = 15$ genetic variants. Data was generated for $2N$ participants, and the associations of the variants with the risk factor were estimated in the first N participants, and associations with the outcome in the second N participants. Only the summary level data (beta-coefficients and standard errors) was used in the analyses. A one-sample setting was also considered where an additional N participants were simulated and all of the genetic associations were estimated from same the N participants.

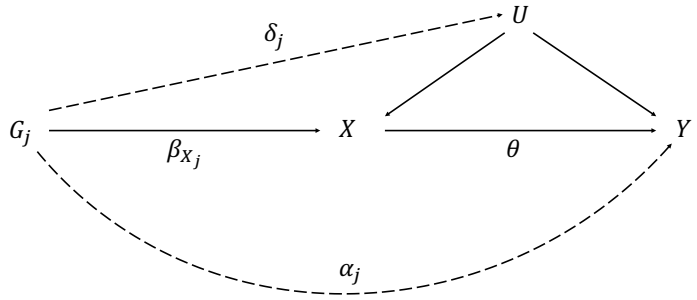


Fig. 3.4 Directed acyclic graph used in the data generating model for the simulation study. U represents the set of unmeasured variables that confound the association between the risk factor X and outcome Y. The genetic effect of G_j on X is β_{X_j}, the direct (pleiotropic) effect of G_j on Y is α_j, the effect of G_j on U is δ_j, and the causal effect of X on Y is θ.

If a genetic variant is associated with a confounder of $X - Y$ association, then this will affect the variant's association with both the risk factor and the outcome, leading to the violation of the InSIDE assumption. Using this observation, data was simulated to consider the following four scenarios:

- Scenario 1 - No pleiotropy, InSIDE automatically satisfied: α_j and δ_j were set to zero for all j.
- Scenario 2 - Balanced pleiotropy, InSIDE satisfied: $\alpha_j \sim U[0.05, 0.15]$ for invalid variants, with each α_j having a 0.5 probability of being multiplied by -1. δ_j was set to zero for all j.
- Scenario 3 - Directional pleiotropy, InSIDE satisfied: $\alpha_j \sim U[0.05, 0.15]$ for invalid variants, and δ_j was set to zero for all j.
- Scenario 4 - Directional pleiotropy, InSIDE violated: $\delta_j \sim U[0.05, 0.10]$ for invalid variants, and α_j was set to zero for all j.

The bounds of the uniform distribution for α_j and δ_j were chosen to ensure the pleiotropic effects were sufficiently large with a good range. To mimic common genetic variants, G_j were coded as 0, 1 or 2, and were generated independently from a Binomial distribution with minor allele frequency 0.3. If a genetic variant was a valid IV then α_j and δ_j were set to zero in all four scenarios. In Scenarios 2 to 4, the number of invalid IVs was set to 1, 3 and 6. The causal effect of the risk factor on the outcome was either $\theta = 0$ (null causal effect) or $\theta = 0.3$ (positive causal effect). To ensure that the pleiotropic effects (α_j and δ_j) and the effects of the genetic variants on the risk factor (β_{X_j}) were comparable, and the amount of variance explained in X by the

genetic variants was approximately 3% across the four scenarios, β_{X_j} was drawn from a uniform distribution between 0.06 and 0.13. 10 000 simulated datasets were generated for each combination of parameters (24 different combinations in total).

3.5.2 Methods applied to the simulated data

We applied the methods discussed in Section 3.3 (summarised in Table 3.1) to the simulated datasets, including the IVW method with: 1) robust regression; 2) penalized weights; 3) robust regression and penalized weights; 4) the three sets of genetic variants from the LTS selection method as outlined in Section 3.3.4; and 5) the genetic variants from the Lasso selection method with the heterogeneity stopping rule. We also applied robust regression, penalized weights, and robust regression and penalized weights to the MR-Egger method. Hence, all of the methods in Table 3.1 were considered in the simulation study apart from Lasso selection with cross validation. The abbreviations in Table 3.1 are used in the result tables for the simulation study.

The bias and coverage properties of the estimates from the robust methods were compared to those from the IVW (with all J genetic variants), simple (unweighted) median, weighted median, and MR-Egger methods. Standard errors for the simple and weighted median estimates were obtained through bootstrapping [52].

The Lasso selection method was applied to $\lambda = 0.1, 0.2, \ldots, 4.9, 5.0, 5.2, 5.4, \ldots, 9.8, 10.0$ under the heterogeneity stopping rule. The first set of variants used in the IVW model under LTS selection was based on the h (approximately 50% of the data) variants used to estimate the initial value of the LTS regression estimate $\hat{\theta}_{LTS,h}$. Since the simulation study generated data for 15 genetic variants, $h = 8$ variants were selected by the LTS objective function for each of the simulated datasets, and these variants were then included in the IVW model. Under the automated approach for LTS selection, h took an initial value of 8 and increased in increments of one up to 15. The number of genetic variants whose weights were either penalized (penalized weights method) or were not selected for the IVW method(Lasso selection and LTS selection methods) were recorded for each simulated dataset.

The maximum number of invalid instruments was set to 6 in the data generating model (Section 3.5.1) to ensure 60% or more of the genetic variants were valid IVs. We anticipated that robust regression and LTS selection would produce consistent estimates of the causal effect for scenarios 1 to 3 as these methods have a breakdown point of 50% in the robust statistics literature. Although we anticipated that the estimates from penalized weights and Lasso selection would be less biased when there was a smaller proportion of invalid IVs, we were unsure whether these methods were

guaranteed to produce consistent estimates if more than 50% of the variants were valid IVs. Since there was only a maximum of 40% invalid IVs, the simple and weighted median methods should produce consistent estimates for scenarios 1 to 3.

3.5.3 Results

The mean proportion of variance in the risk factor explained by the genetic variants (R^2 statistic), mean F statistic, and mean I^2 statistic are contained in Table 3.3 for scenarios 1-4 for the null and positive causal effects by the number of invalid instruments. The mean R^2 values were greater than 3% for all of the scenarios, and the minimum mean F-statistic was 20.8. The I^2 statistic ranged from 39.1% to 77.5%. Since violations in the NOME assumption can lead to attenuation towards the null for the MR-Egger estimates, and this attenuation is approximately equal to the I^2 statistic, we expected the MR-Egger estimates for the positive causal effect to be severely attenuated towards the null [49].

Table 3.3 Mean values of the R^2 (%), F-statistic and I^2 (%) for Scenarios 1-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instrumental variables (IV).

	No invalid IVs			1 invalid IV			3 invalid IVs			6 invalid IVs		
	R^2	F	I^2	R^2	F	I^2	R^2	F	I^2	R^2	F	I^2
Null causal effect: $\theta = 0$												
Scenario 1	3.0	20.8	39.6	-	-	-	-	-	-	-	-	
Scenario 2	-	-	-	3.0	20.8	39.6	3.0	20.8	39.3	3.0	20.8	39.5
Scenario 3	-	-	-	3.0	20.8	39.7	3.0	20.8	39.5	3.0	20.8	39.2
Scenario 4	-	-	-	3.4	23.6	56.5	4.2	29.3	70.7	5.4	37.7	77.5
Positive causal effect: $\theta = 0.3$												
Scenario 1	3.0	20.8	39.3	-	-	-	-	-	-	-	-	
Scenario 2	-	-	-	3.0	20.8	39.1	3.0	20.8	39.4	3.0	20.8	39.6
Scenario 3	-	-	-	3.0	20.8	39.9	3.0	20.8	39.7	3.0	20.8	39.6
Scenario 4	-	-	-	3.4	23.6	56.4	4.2	29.3	70.8	5.4	37.7	77.4

Results from the simulation study for the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (Rr); 3) penalized weights (PW); 4) robust regression and penalized weights; and 5) the three sets of variants selected by LTS selection are provided in Table 3.4 (Scenario 1 only), Table 3.5 (null causal effect $\theta = 0$), and Table 3.6 (positive causal effect $\theta = 0.3$). Tables 3.4 to 3.6 also contain the results from Lasso selection with the heterogeneity stopping rule, simple (unweighted)

median, weighted median and MR-Egger methods, and for each method, information on the: mean estimate; mean standard error of the estimates; standard deviation of the estimates; coverage of the true causal effect of the 95% confidence interval; and power to detect the causal effect at the 5% significance level are provided. The power (at the 5% significance level) of the intercept test in the MR-Egger method for detecting directional pleiotropy and/or violation of the InSIDE assumption in all scenarios is provided in Table 3.7. The number of robust regression models that did not report a standard error (maximum of 2.6% across all of the scenarios considered) are given in the Table H.1. Apart from the calculation of the mean standard error, the robust regression models that did not report a standard error were included in the results, and the power calculations treated the standard error as infinite.

When all of the genetic variants were valid IVs (Table 3.4), all of the methods produced unbiased estimates of the null causal effect. With the exception of the IVW model with the h variants selected from LTS selection, the Type 1 error rates for the null causal effect were close to the nominal level of 5%. Apart from the simple median method, there was attenuation towards the null with a positive causal effect for all methods, and as expected, this was particularly evident for the MR-Egger method (also observed for Scenarios 2 and 3). Violation of the NOME assumption can lead to inflation of the intercept term in the MR-Egger method [49], and this was true for the simulation study where the power to detect the intercept term for Scenarios 1 and 2 was greater than 5% (Table 3.7). Only 7.5% of the MR-Egger models detected a positive causal effect, and apart from the median estimators and the IVW model with the h variants from LTS, all of the robust methods had approximately 95% power to detect the positive causal effect.

Although the mean estimates in Scenario 2 (Tables 3.5 and 3.6) were similar to those in Scenario 1, there were clear differences in the precision of the estimates for the null and positive causal effects, with most of the methods reporting larger mean standard errors under Scenario 2. With the exception of the IVW model with the h variants from LTS selection, where the mean standard error remained constant, the mean standard error increased as the number of invalid instruments increased for all methods. As seen in Scenario 1, the IVW model with the h variants from LTS selection had inflated Type I error rates and poor coverage. The IVW model with penalized weights had the most precise estimates, but suffered from inflated Type I error rates and poor coverage. The estimates from Lasso selection and the automated approach to LTS selection were almost identical for Scenarios 1 and 2. The simple and weighted

Table 3.4 Mean (standard error), standard deviation, coverage (%), and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression; 3) penalized weights; 4) robust regression and penalized weights; 5) the three sets of variants selected by the least trimmed squared (LTS) estimator; and 6) the genetic variants from the Lasso selection method with the heterogeneity stopping rule for Scenario 1 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect. Results from the simple (unweighted) median, weighted median and MR-Egger methods are also provided.

	Null causal effect ($\theta = 0$)				**Positive causal effect** ($\theta = 0.3$)			
	Estimate (SE)	SD	Cov.	Pow.	Estimate (SE)	SD	Cov.	Pow.
Scenario 1. No pleiotropy, InSIDE automatically satisfied								
IVW	-0.001 (0.061)	0.058	95.7	4.3	0.287 (0.073)	0.069	95.5	98.2
Robust regression	-0.001 (0.066)	0.060	95.1	4.9	0.287 (0.079)	0.072	94.7	94.8
Penalized weights	-0.001 (0.060)	0.059	95.0	5.0	0.289 (0.072)	0.071	94.7	98.2
Robust regression with penalized weights	-0.001 (0.064)	0.061	94.5	5.5	0.288 (0.077)	0.073	94.1	95.7
LTS[a]								
Variants from h	-0.001 (0.078)	0.116	81.9	18.1	0.291 (0.092)	0.140	80.4	78.5
Variants from $w_{LTS,2}$	-0.001 (0.061)	0.064	93.3	6.7	0.289 (0.073)	0.077	93.1	97.4
Automated approach	-0.001 (0.060)	0.059	95.1	4.9	0.287 (0.072)	0.071	94.9	98.1
Lasso selection	-0.001 (0.060)	0.059	94.8	5.2	0.287 (0.072)	0.071	94.6	98.0
Median								
Simple	-0.002 (0.086)	0.074	97.9	2.1	0.301 (0.105)	0.090	98.0	86.9
Weighted	-0.002 (0.080)	0.071	97.4	2.6	0.277 (0.097)	0.085	96.7	85.7
MR-Egger	-0.001 (0.219)	0.207	96.1	3.9	0.143 (0.261)	0.251	91.0	7.5

Abbreviations: SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse-variance weighted; LTS, least trimmed squares.
[a]The following three sets of genetic variants were selected from the LTS estimator and included in the IVW model: 1) the h=8 variants used to estimate the initial LTS estimate $\hat{\theta}_{LTS,h}$; 2) the variants with a weight of 1 in $w_{LTS,2}$; and 3) the variants selected from the automated approach based on the heterogeneity stopping rule.

median estimators performed just as well, if not better, than the other robust methods for Scenario 2.

In Scenario 3 (directional pleiotropy, InSIDE satisfied), the IVW method produced biased causal estimates with inflated Type 1 error rates, and the degree of bias increased with the number of invalid IVs. With one invalid instrument, estimates from the robust methods were slightly biased and Type 1 error rates were fairly well controlled (with the exception of the IVW model with the h variants from LTS selection). As the number of instruments increased, bias in the estimates for the robust methods also increased, although the magnitude of bias was smaller than the IVW method, and Type 1 error inflation was less severe. The performance of the LTS selection method varied, and the estimates based on the h variants were the least biased across all of the robust methods, however, as with Scenarios 1 and 2, the estimates were too precise and had poor coverage. The estimates from the LTS selection method using the automated approach and Lasso selection produced almost identical results with the exception of 6

invalid instruments, where the LTS selection method produced less biased estimates. Robust regression with penalized weights performed reasonably well when there was 1 or 3 invalid instruments.

In Scenario 4 (directional pleiotropy, InSIDE violated), all of the robust methods produced biased estimates. When there was only one invalid instrument, the magnitude of bias from the robust methods was less severe than the IVW method, and this was particularly true for robust regression with penalized weights and LTS selection. Unlike robust regression with penalized weights, LTS selection suffered from poor coverage of the causal effect when there was one invalid IV. As the number of invalid IVs increased, the performance of the robust methods worsened, and there was little advantage in applying the robust methods compared to the median estimator in Scenario 4 when 6 of the 15 genetic variants were invalid IVs.

Table 3.5 Mean (standard error), standard deviation, coverage (%), and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (Rr); 3) penalized weights (PW); 4) robust regression and penalized weights (Rr and PW); 5) the three sets of variants selected by the least trimmed squared (LTS) estimator; and 6) the genetic variants from the Lasso selection (LS) method with the heterogeneity stopping rule for Scenarios 2-4 with a null causal effect ($\theta = 0$) by the number of invalid IVs. Results from the simple median, weighted median and MR-Egger methods are also provided.

	1 invalid IV				**3 invalid IVs**				**6 invalid IVs**			
	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.
Scenario 2. Balanced pleiotropy, InSIDE satisfied												
IVW	-0.002 (0.089)	0.092	94.7	5.3	0.000 (0.133)	0.136	93.4	6.6	0.000 (0.180)	0.183	93.0	7.0
Rr	-0.002 (0.069)	0.065	94.3	5.7	0.000 (0.096)	0.087	94.5	5.5	0.001 (0.196)	0.173	94.3	5.6
PW	-0.002 (0.062)	0.064	94.2	5.8	0.000 (0.066)	0.077	91.1	8.9	0.001 (0.075)	0.116	81.5	18.5
Rr and PW	-0.002 (0.071)	0.065	94.6	5.4	0.001 (0.094)	0.078	94.7	5.2	0.001 (0.160)	0.119	91.5	7.3
LTS[a]												
h	-0.002 (0.078)	0.115	82.2	17.8	0.001 (0.079)	0.113	83.7	16.3	0.001 (0.080)	0.118	86.2	13.8
$w_{LTS,2}$	-0.001 (0.064)	0.066	94.2	5.8	0.001 (0.075)	0.086	92.0	8.0	0.002 (0.142)	0.180	85.2	14.8
Auto	-0.002 (0.064)	0.064	94.6	5.4	0.000 (0.071)	0.081	91.8	8.3	0.001 (0.091)	0.136	83.5	16.5
LS	-0.002 (0.063)	0.065	94.4	5.6	0.000 (0.071)	0.080	91.7	8.3	0.001 (0.088)	0.129	84.5	15.5
Median												
Simple	-0.002 (0.090)	0.080	97.4	2.6	0.001 (0.097)	0.094	96.5	3.5	0.002 (0.115)	0.132	92.7	7.3
Weighted	-0.001 (0.082)	0.076	96.9	3.2	0.000 (0.089)	0.090	95.2	4.8	0.000 (0.101)	0.133	88.9	11.1
MR-Egger	-0.004 (0.317)	0.335	92.7	7.3	-0.009 (0.477)	0.496	92.7	7.3	-0.006 (0.646)	0.661	93.0	7.0
Scenario 3. Directional pleiotropy, InSIDE satisfied												
IVW	0.064 (0.089)	0.064	94.8	5.2	0.194 (0.126)	0.076	76.0	24.0	0.388 (0.154)	0.089	16.1	83.9
Rr	0.010 (0.069)	0.064	94.3	5.7	0.069 (0.113)	0.083	93.9	6.1	0.335 (0.227)	0.105	63.6	36.4
PW	0.007 (0.062)	0.063	94.2	5.8	0.033 (0.067)	0.078	89.2	10.8	0.148 (0.082)	0.137	57.3	42.7
Rr and PW	0.005 (0.072)	0.065	94.8	5.2	0.025 (0.092)	0.079	93.2	6.7	0.115 (0.138)	0.147	78.6	20.9
LTS[a]												
h	0.000 (0.078)	0.115	82.6	17.4	0.007 (0.079)	0.116	83.0	17.0	0.030 (0.080)	0.143	84.1	15.9
$w_{LTS,2}$	0.004 (0.064)	0.066	94.2	5.8	0.041 (0.076)	0.085	90.5	9.5	0.283 (0.130)	0.155	37.6	62.4
Auto	0.006 (0.063)	0.064	94.5	5.5	0.030 (0.071)	0.081	90.5	9.5	0.119 (0.093)	0.160	68.7	31.3
LS	0.006 (0.063)	0.065	94.2	5.8	0.031 (0.071)	0.080	90.3	9.7	0.164 (0.096)	0.146	60.5	39.5
Median												
Simple	0.021 (0.089)	0.077	97.3	2.7	0.074 (0.100)	0.086	92.9	7.2	0.224 (0.134)	0.124	64.3	35.7
Weighted	0.017 (0.082)	0.074	96.9	3.1	0.065 (0.090)	0.085	91.7	8.3	0.210 (0.110)	0.149	56.9	43.1
MR-Egger	-0.003 (0.318)	0.334	92.9	7.2	-0.001 (0.450)	0.465	93.2	6.8	-0.004 (0.544)	0.562	92.4	7.6
Scenario 4. Directional pleiotropy, InSIDE violated												
IVW	0.077 (0.070)	0.058	83.4	16.7	0.186 (0.075)	0.056	25.6	74.4	0.290 (0.071)	0.050	0.3	99.7
Rr	0.031 (0.085)	0.069	93.9	6.0	0.142 (0.127)	0.082	73.1	26.0	0.289 (0.079)	0.053	3.3	96.6
PW	0.021 (0.061)	0.070	89.2	10.8	0.083 (0.063)	0.091	64.1	35.9	0.231 (0.061)	0.092	12.2	87.8
Rr and PW	0.018 (0.071)	0.070	92.7	7.3	0.075 (0.084)	0.095	76.2	23.7	0.230 (0.074)	0.101	19.0	80.8
LTS[a]												
h	0.005 (0.078)	0.121	79.6	20.4	0.030 (0.076)	0.139	75.0	25.0	0.162 (0.066)	0.197	45.7	54.3
$w_{LTS,2}$	0.017 (0.063)	0.074	89.7	10.3	0.090 (0.067)	0.103	61.5	38.5	0.266 (0.069)	0.088	5.8	94.3
Auto	0.025 (0.062)	0.073	88.1	11.9	0.105 (0.066)	0.108	51.9	48.1	0.265 (0.067)	0.103	6.7	93.3
LS	0.024 (0.062)	0.073	88.2	11.8	0.116 (0.066)	0.099	51.1	48.9	0.286 (0.066)	0.070	2.2	97.8
Median												
Simple	0.020 (0.089)	0.077	97.3	2.7	0.071 (0.092)	0.083	89.9	10.1	0.192 (0.088)	0.091	40.0	60.0
Weighted	0.055 (0.082)	0.077	91.0	9.0	0.198 (0.081)	0.097	34.8	65.2	0.343 (0.069)	0.074	0.5	99.5
MR-Egger	0.305 (0.214)	0.219	66.8	33.2	0.539 (0.197)	0.183	21.3	78.7	0.644 (0.182)	0.165	5.1	94.9

Abbreviations: IV, instrumental variable; Est. estimate; SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted; Rr, robust regression; PW, penalized weights; LTS, least trimmed squares; LS, Lasso selection; Auto, automated.

[a]The following three sets of genetic variants were selected from the LTS estimator and included in the IVW model: 1) the h=8 variants used to estimate the initial LTS estimate $\hat{\theta}_{LTS,h}$; 2) the variants with a weight of 1 in $w_{LTS,2}$; and 3) the variants selected from the automated approach based on the heterogeneity stopping rule.

Table 3.6 Mean (standard error), standard deviation, coverage (%), and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (Rr); 3) penalized weights (PW); 4) robust regression and penalized weights (Rr and PW); 5) the three sets of variants selected by the least trimmed squared (LTS) estimator; and 6) the genetic variants from the Lasso selection (LS) method with the heterogeneity stopping rule for Scenarios 2-4 with a positive causal effect ($\theta = 0.3$) by the number of invalid IVs. Results from the simple median, weighted median and MR-Egger methods are also provided.

	1 invalid IV				**3 invalid IVs**				**6 invalid IVs**			
	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.
Scenario 2. Balanced pleiotropy, InSIDE satisfied												
IVW	0.286 (0.097)	0.100	94.0	80.9	0.288 (0.139)	0.140	93.5	54.5	0.285 (0.184)	0.184	93.3	34.3
Rr	0.287 (0.084)	0.079	93.9	91.0	0.288 (0.116)	0.107	93.6	71.1	0.286 (0.193)	0.178	93.8	34.5
PW	0.289 (0.074)	0.079	93.2	96.5	0.291 (0.080)	0.098	88.6	91.8	0.295 (0.090)	0.147	78.4	80.5
Rr and PW	0.289 (0.083)	0.080	93.5	92.5	0.290 (0.100)	0.097	92.2	81.6	0.295 (0.145)	0.147	88.1	59.4
LTS[a]												
h	0.292 (0.092)	0.141	80.2	78.8	0.292 (0.093)	0.143	80.5	78.9	0.292 (0.094)	0.164	79.2	78.5
$w_{LTS,2}$	0.288 (0.076)	0.081	93.2	95.1	0.289 (0.092)	0.108	90.2	83.6	0.287 (0.156)	0.192	85.4	46.5
Auto	0.287 (0.076)	0.079	93.5	95.3	0.289 (0.086)	0.103	89.6	87.2	0.289 (0.112)	0.178	79.0	67.1
LS	0.287 (0.076)	0.080	93.2	95.4	0.288 (0.085)	0.102	89.4	88.0	0.288 (0.108)	0.167	80.5	69.9
Median												
Simple	0.302 (0.109)	0.097	97.5	83.1	0.302 (0.118)	0.113	96.1	75.3	0.303 (0.136)	0.155	92.5	61.4
Weighted	0.276 (0.100)	0.091	96.1	81.9	0.277 (0.106)	0.108	94.0	75.6	0.277 (0.119)	0.152	88.2	63.2
MR-Egger	0.143 (0.349)	0.363	90.2	9.0	0.138 (0.495)	0.518	91.2	8.4	0.126 (0.657)	0.681	92.1	7.5
Scenario 3. Directional pleiotropy, InSIDE satisfied												
IVW	0.353 (0.098)	0.075	96.1	97.8	0.482 (0.133)	0.087	81.5	99.3	0.673 (0.160)	0.101	28.3	100
Rr	0.306 (0.084)	0.077	95.1	94.8	0.383 (0.134)	0.099	93.9	86.2	0.631 (0.205)	0.112	60.8	90.8
PW	0.303 (0.074)	0.078	93.8	98.0	0.346 (0.081)	0.100	86.4	97.8	0.511 (0.098)	0.164	47.6	99.1
Rr and PW	0.300 (0.083)	0.080	94.1	93.5	0.335 (0.102)	0.102	91.3	88.8	0.485 (0.142)	0.179	66.2	86.9
LTS[a]												
h	0.297 (0.092)	0.144	79.3	79.2	0.308 (0.093)	0.148	80.2	81.4	0.366 (0.094)	0.214	75.1	86.4
$w_{LTS,2}$	0.299 (0.076)	0.081	93.4	97.0	0.355 (0.092)	0.107	88.8	96.8	0.605 (0.142)	0.147	39.6	99.2
Auto	0.301 (0.076)	0.078	94.3	97.5	0.340 (0.087)	0.105	88.4	96.3	0.490 (0.114)	0.188	56.2	95.9
LS	0.301 (0.076)	0.079	94.0	97.4	0.340 (0.086)	0.104	88.2	96.5	0.513 (0.113)	0.168	53.8	98.5
Median												
Simple	0.329 (0.110)	0.095	97.5	89.5	0.393 (0.125)	0.108	93.5	93.2	0.572 (0.158)	0.150	61.8	97.2
Weighted	0.300 (0.100)	0.090	97.2	88.1	0.356 (0.111)	0.104	94.5	93.0	0.516 (0.131)	0.161	63.6	97.4
MR-Egger	0.142 (0.345)	0.353	90.9	8.7	0.138 (0.468)	0.485	91.3	8.1	0.137 (0.555)	0.576	91.9	8.0
Scenario 4. Directional pleiotropy, InSIDE violated												
IVW	0.367 (0.080)	0.071	88.8	99.8	0.478 (0.084)	0.067	42.2	100	0.582 (0.078)	0.062	2.3	100
Rr	0.329 (0.100)	0.082	94.3	90.2	0.447 (0.128)	0.085	73.7	90.2	0.581 (0.087)	0.066	8.2	99.7
PW	0.323 (0.072)	0.086	88.6	98.2	0.403 (0.072)	0.102	61.2	98.9	0.546 (0.068)	0.092	11.2	99.9
Rr and PW	0.318 (0.085)	0.087	92.4	94.1	0.397 (0.095)	0.107	72.7	93.6	0.547 (0.077)	0.098	16.0	98.4
LTS[a]												
h	0.306 (0.091)	0.148	77.1	80.9	0.345 (0.087)	0.175	67.5	83.2	0.492 (0.075)	0.205	37.8	93.0
$w_{LTS,2}$	0.316 (0.074)	0.091	88.2	97.1	0.406 (0.077)	0.112	60.9	98.3	0.567 (0.075)	0.089	6.8	99.6
Auto	0.324 (0.074)	0.088	88.1	97.8	0.424 (0.076)	0.112	52.5	98.4	0.570 (0.074)	0.094	5.4	99.2
LS	0.323 (0.073)	0.089	87.9	97.6	0.430 (0.076)	0.105	52.7	99.3	0.579 (0.073)	0.079	4.4	100
Median												
Simple	0.328 (0.108)	0.095	97.0	89.9	0.387 (0.111)	0.101	90.0	95.1	0.509 (0.101)	0.101	43.4	99.5
Weighted	0.344 (0.099)	0.095	94.0	94.4	0.496 (0.097)	0.108	47.2	99.7	0.625 (0.085)	0.087	3.4	100
MR-Egger	0.488 (0.254)	0.259	86.1	51.8	0.767 (0.233)	0.220	45.4	90.0	0.887 (0.214)	0.197	20.2	98.1

Abbreviations: IV, instrumental variable; Est. estimate; SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted; Rr, robust regression; PW, penalized weights; LTS, least trimmed squares; LS, Lasso selection; Auto, automated.

[a]The following three sets of genetic variants were selected from the LTS estimator and included in the IVW model: 1) the h=8 variants used to estimate the initial LTS estimate $\hat{\theta}_{LTS,h}$; 2) the variants with a weight of 1 in $w_{LTS,2}$; and 3) the variants selected from the automated approach based on the heterogeneity stopping rule.

Table 3.7 Power (%) of the intercept test in the MR-Egger method for detecting directional pleiotropy and/or violation of the InSIDE assumption for Scenarios 1-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instrumental variables (IV).

	Null causal effect				Positive causal effect			
No. invalid:	0	1	3	6	0	1	3	6
Scenario 1	3.7	-	-	-	8.7	-	-	
Scenario 2	-	7.2	7.5	7.0	-	9.4	8.5	7.8
Scenario 3	-	7.2	8.7	13.1	-	11.2	13.8	19.1
Scenario 4	-	8.6	26.2	32.0	-	22.8	49.9	55.9

The performance of the penalized weights, Lasso selection and LTS selection methods can also be evaluated by considering the mean number of genetic variants whose weights were penalized or were not selected for the IVW model (Table 3.8). With the exception of the scenario when there was only one invalid instrument, the mean numbers of penalized or not selected variants were noticeably smaller than the actual number of invalid instruments for all of the robust methods. There was little difference between the mean number of variants penalized or not selected for Scenarios 2 and 3 for the different methods. However, there were large reductions in the mean number of variants penalized or not selected for the IVW method for Scenario 4 compared to Scenarios 2 and 3. As the number of invalid IVs increased from 1 to 6, the percentage of simulated datasets that correctly penalized or did not include all of the invalid instruments decreased considerably. In terms of the mean number of variants penalized or not selected for the IVW method, and the frequency that all invalid instruments were correctly downweighted or not selected, the IVW method with penalized weights was generally the most effective method across the different scenarios for both the null and positive causal effects.

Results from applying robust regression and penalized weights to the MR-Egger method are provided in Table H.2. We had hoped that combining these robust methods with the MR-Egger method would provide additional robustness. In particular, we anticipated that there would be less bias in the estimates of the causal effect when the robust methods and MR-Egger were combined. However, the results were disappointing as there was no improvement in the performance of the methods when they were combined, and all of the models were affected by the violation of the NOME assumption.

Finally, results from the one-sample setting are provided in the Table H.3 and Table H.4. Bias in the direction of the observational association was observed in all methods. As with the two–sample setting, the IVW model with the h variants from

LTS selection produced the least biased estimates for all scenarios, and the IVW with penalized weights was the most precise.

Table 3.8 Mean estimate, mean number (minimum, maximum) and standard deviation of variants penalized or not selected for the IVW method, and the frequency (%) all invalid instruments had their weights penalized by the penalized weights method or were not selected for the IVW method under Lasso selection with the heterogeneity stopping rule or LTS selection for $w_{LTS,2}$ and the automated approach for Scenarios 2-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instruments.

	Null causal effect ($\theta = 0$)				**Positive causal effect** ($\theta = 0.3$)			
	Mean estimate	Mean no. (min, max)	SD	Freq. (%)	Mean estimate	Mean no. (min, max)	SD	Freq. (%)
1 invalid IV								
Scenario 2. Balanced pleiotropy, InSIDE satisfied								
Penalized weights	-0.002	1.04 (0, 5)	0.542	86.2	0.289	1.01 (0, 5)	0.632	78.0
Lasso selection	-0.002	0.91 (0, 8)	0.621	80.3	0.287	0.83 (0, 8)	0.730	69.8
LTS $w_{LTS,2}$ variants	-0.001	1.10 (0, 7)	0.739	85.3	0.288	1.01 (0, 7)	0.804	75.2
LTS automated	-0.002	0.84 (0, 5)	0.475	79.3	0.287	0.74 (0, 5)	0.544	68.0
Scenario 3. Directional pleiotropy, InSIDE satisfied								
Penalized weights	0.007	1.04 (0, 5)	0.540	86.5	0.303	1.00 (0, 5)	0.629	76.8
Lasso selection	0.006	0.91 (0, 10)	0.623	80.7	0.301	0.81 (0, 9)	0.681	69.7
LTS $w_{LTS,2}$ variants	0.004	1.12 (0, 7)	0.737	85.7	0.299	1.02 (0, 7)	0.786	75.4
LTS automated	0.006	0.85 (0, 4)	0.467	79.9	0.301	0.74 (0, 5)	0.536	68.4
Scenario 4. Directional pleiotropy, InSIDE violated								
Penalized weights	0.021	0.88 (0, 5)	0.614	70.1	0.301	0.77 (0, 4)	0.673	54.7
Lasso selection	0.024	0.69 (0, 9)	0.725	56.2	0.301	0.57 (0, 10)	0.764	42.3
LTS $w_{LTS,2}$ variants	0.017	0.95 (0, 7)	0.835	67.7	0.316	0.85 (0, 7)	0.878	54.2
LTS automated	0.025	0.59 (0, 3)	0.558	54	0.324	0.46 (0, 3)	0.559	40.0
3 invalid IV								
Scenario 2. Balanced pleiotropy, InSIDE satisfied								
Penalized weights	0.000	2.73 (0, 6)	0.723	62.4	0.291	2.51 (0, 7)	0.853	45.0
Lasso selection	0.000	2.48 (0, 10)	0.928	52.5	0.288	2.16 (0, 9)	1.10	35.8
LTS $w_{LTS,2}$ variants	0.001	2.21 (0, 7)	1.000	41.7	0.289	1.78 (0, 7)	1.07	24.6
LTS automated	0.000	2.35 (0, 6)	0.837	49.5	0.289	1.98 (0, 6)	0.96	31.8
Scenario 3. Directional pleiotropy, InSIDE satisfied								
Penalized weights	0.033	2.67 (0, 6)	0.790	56.1	0.346	2.46 (0, 7)	0.942	38.6
Lasso selection	0.031	2.48 (0, 9)	1.06	52.4	0.340	2.14 (0, 10)	1.21	36.1
LTS $w_{LTS,2}$ variants	0.041	2.15 (0, 7)	1.02	41.2	0.355	1.75 (0, 7)	1.13	26.1
LTS automated	0.030	2.29 (0, 5)	0.912	49.5	0.340	1.92 (0, 7)	1.04	32.9
Scenario 4. Directional pleiotropy, InSIDE violated								
Penalized weights	0.083	1.97 (0, 6)	1.01	26.2	0.336	1.58 (0, 5)	1.04	12.8
Lasso selection	0.116	1.45 (0, 11)	1.68	16.4	0.348	1.08 (0, 11)	1.50	8.2
LTS $w_{LTS,2}$ variants	0.090	1.41 (0, 7)	1.27	22.1	0.406	1.10 (0, 7)	1.19	12.4
LTS automated	0.105	1.10 (0, 5)	1.19	18.9	0.424	0.74 (0, 6)	1.03	9.1
6 invalid IV								
Scenario 2. Balanced pleiotropy, InSIDE satisfied								
Penalized weights	0.001	5.26 (1, 9)	0.979	37.4	0.295	4.73 (0, 9)	1.14	18.5
Lasso selection	0.001	4.72 (0, 12)	1.64	31.8	0.288	3.87 (0, 13)	1.86	15.5
LTS $w_{LTS,2}$ variants	0.002	1.64 (0, 7)	1.8	5.3	0.287	1.12 (0, 7)	1.48	1.8
LTS automated	0.001	4.28 (0, 7)	1.53	24.6	0.289	3.30 (0, 7)	1.67	9.5
Scenario 3. Directional pleiotropy, InSIDE satisfied								
Penalized weights	0.148	4.94 (0, 11)	1.36	15.6	0.511	4.23 (0, 10)	1.45	5.1
Lasso selection	0.164	4.63 (0, 13)	2.75	23.3	0.513	3.37 (0, 13)	2.65	10.5
LTS $w_{LTS,2}$ variants	0.283	1.40 (0, 7)	1.82	5.7	0.605	0.93 (0, 7)	1.50	2.5
LTS automated	0.119	3.78 (0, 7)	2.02	26.2	0.490	2.60 (0, 7)	2.09	12.2
Scenario 4. Directional pleiotropy, InSIDE violated								
Penalized weights	0.231	2.36 (0, 7)	1.48	1.6	0.399	1.81 (0, 7)	1.30	0.4
Lasso selection	0.286	0.97 (0, 13)	1.87	0.1	0.446	0.83 (0, 13)	1.58	0.0
LTS $w_{LTS,2}$ variants	0.266	0.59 (0, 7)	1.29	2.2	0.448	0.52 (0, 7)	1.08	0.9
LTS automated	0.265	0.63 (0, 7)	1.44	3.8	0.570	0.44 (0, 7)	1.03	1.2

Abbreviations: no., number; min, minimum; max, maximum; SD, standard deviation; Freq., frequency; IV, instrumental variable; InSIDE, instrument strength independent of direct effect; LTS, least trimmed squares.

3.5.4 Increased number of genetic variants

Since many of the methods outlined in Section 3.3 are based on asymptotic theory, it was anticipated that there would be an improvement in the performance of the methods when the data was generated with a larger number of genetic variants. We therefore repeated the simulation study outlined in Section 3.5.1 for Scenarios 2–4 for 1 000 simulated datasets with the number of genetic variants increased from 15 to 100, and the number of invalid IVs was multiplied by 5, increasing from 1, 3 and 6 to 5, 15 and 30. The bounds of the uniform distribution used to generate the genetic associations with the risk factor (β_{X_j}) were divided by $\frac{\sqrt{100}}{\sqrt{15}}$ to ensure the average R^2 values were comparable with the original simulation study. The IVW model with: 1) the full set of genetic variants; 2) robust regression; 3) penalized weights; 4) robust regression and penalized weights; 5) the h genetic variants obtained from the LTS selection method (Section 3.3.4); and 6) the genetic variants from the Lasso selection method with the heterogeneity stopping rule were all applied to the dataset.

Results

The mean R^2 statistic, F-statistic, and I^2 statistic are contained in Table 3.9 for Scenarios 2–4 for the null and positive causal effect by the number of invalid IVs. The mean R^2 values for the 100 genetic variants were slightly higher than the values reported in the original simulation study (Table 3.3). For all of the scenarios considered, there was a significant reduction in the mean F-statistic and I^2 statistic, and we therefore expected the estimates to be affected by weak instrument bias.

Table 3.9 Mean values of the R^2 (%), F-statistic and I^2 (%) for Scenarios 2-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instrumental variables (IV) when the simulation study was re-performed for 100 genetic variants.

	5 invalid IV			15 invalid IVs			30 invalid IVs		
	R^2	F	I^2	R^2	F	I^2	R^2	F	I^2
Null causal effect: $\boldsymbol{\theta = 0}$									
Scenario 2	4.0	4.2	3.0	4.0	4.2	3.3	4.0	4.2	2.9
Scenario 3	4.0	4.2	3.1	4.0	4.2	3.0	4.0	4.2	3.0
Scenario 4	5.2	5.4	32.9	7.3	7.8	58.3	10.3	11.4	69.5
Positive causal effect: $\boldsymbol{\theta = 0.3}$									
Scenario 2	4.0	4.2	3.2	4.0	4.2	3.1	4.1	4.2	3.1
Scenario 3	4.0	4.2	3.3	4.1	4.2	3.1	4.0	4.2	2.9
Scenario 4	5.2	5.4	33.3	7.3	7.8	58.2	10.3	11.4	69.4

Results from the simulation study for the IVW model with: 1) the J genetic variants (IVW); 2) robust regression; 3) penalized weights; 4) robust regression and penalized weights; 5) the h genetic variants selected by the least trimmed squared (LTS) estimator; and 6) the genetic variants from the Lasso selection method with the heterogeneity stopping rule are provided in Table 3.10 (null causal effect $\theta = 0$), and Table 3.11 (positive causal effect $\theta = 0.3$).

The reduction in the strength of the IVs led to weak instrument bias, and there was severe attenuation towards the null for the positive causal effect (Table 3.11). For the null causal effect, there was little difference in the performance of the robust methods with the increased number of genetic variants. In fact, the methods performed worst under Scenario 4 when 100 variants were included in the data generating model rather than 15 (Table 3.10). Due to the attenuation of the positive causal effect when the number of variants was increased to 100, it was difficult to compare the results to the original simulations. Nevertheless, there was no evidence to suggest that the performances of the robust methods improved when the number of genetic variants was increased.

Table 3.10 Results from the simulation study when 100 genetic variants were simulated for 1 000 datasets. Mean (standard error), standard deviation, coverage (%), and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (Rr); 3) penalized weights (PW); 4) robust regression and penalized weights (Rr and PW); 5) the h genetic variants selected by the least trimmed squared (LTS) estimator; and 6) the genetic variants from the Lasso selection (LS) method with the heterogeneity stopping rule for Scenarios 2-4 with a null causal effect ($\theta = 0$) by the number of invalid instrumental variables.

	5 invalid IV				**15 invalid IVs**				**30 invalid IVs**			
	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.
Scenario 2. Balanced pleiotropy, InSIDE satisfied												
IVW	-0.003 (0.072)	0.071	95.0	5.0	-0.003 (0.103)	0.105	94.9	5.1	0.000 (0.138)	0.144	94.0	6.0
Rr	-0.001 (0.054)	0.051	95.8	4.2	-0.001 (0.065)	0.066	93.7	6.3	0.005 (0.115)	0.114	95.7	4.3
PW	-0.001 (0.051)	0.051	94.8	5.2	-0.001 (0.054)	0.063	91.3	8.7	0.001 (0.060)	0.081	86.3	13.7
Rr and PW	-0.001 (0.055)	0.052	95.8	4.2	-0.001 (0.064)	0.062	95.6	4.4	0.000 (0.087)	0.078	96.5	3.5
LTS[a]												
h	0.001 (0.069)	0.134	69.1	30.9	0.001 (0.070)	0.127	71.8	28.2	-0.003 (0.072)	0.122	75.9	24.1
Auto	-0.001 (0.052)	0.052	94.8	5.2	0.000 (0.057)	0.066	91.6	8.4	0.002 (0.070)	0.095	86.9	13.1
LS	-0.001 (0.051)	0.052	94.2	5.8	-0.002 (0.055)	0.063	91.5	8.5	0.002 (0.062)	0.081	87.6	12.4
Scenario 3. Directional pleiotropy, InSIDE satisfied												
IVW	0.096 (0.071)	0.058	77.3	22.7	0.287 (0.099)	0.070	9.0	91.0	0.572 (0.126)	0.088	0.0	100
Rr	0.014 (0.055)	0.053	94.8	5.2	0.070 (0.072)	0.064	87.0	13.0	0.355 (0.169)	0.103	41.3	58.7
PW	0.012 (0.051)	0.053	93.9	6.1	0.043 (0.054)	0.061	83.9	16.1	0.156 (0.062)	0.094	34.3	65.7
Rr and PW	0.009 (0.055)	0.054	95.2	4.7	0.031 (0.064)	0.061	91.7	8.3	0.108 (0.087)	0.093	74.6	25.4
LTS[a]												
h	0.003 (0.069)	0.138	66.1	33.9	0.000 (0.070)	0.126	73.4	26.6	0.008 (0.071)	0.116	78.6	21.4
Auto	0.012 (0.052)	0.053	94.2	5.8	0.039 (0.057)	0.065	85.6	14.4	0.146 (0.072)	0.118	51.2	48.8
LS	0.013 (0.051)	0.054	93.2	6.8	0.044 (0.055)	0.062	83.4	16.6	0.165 (0.063)	0.095	32.6	67.4
Scenario 4. Directional pleiotropy, InSIDE violated												
IVW	0.170 (0.052)	0.046	7.2	92.8	0.349 (0.049)	0.043	0.0	100	0.476 (0.043)	0.036	0.0	100
Rr	0.076 (0.079)	0.065	87.2	12.8	0.310 (0.089)	0.057	7.7	92.1	0.475 (0.048)	0.038	0.0	100
PW	0.053 (0.049)	0.064	72.9	27.1	0.187 (0.046)	0.082	13.5	86.5	0.401 (0.040)	0.062	0.0	100
Rr and PW	0.047 (0.058)	0.064	84.5	15.5	0.184 (0.065)	0.087	26.7	73.3	0.409 (0.044)	0.062	0.0	100
LTS[a]												
h	0.014 (0.068)	0.136	68.8	31.2	0.042 (0.066)	0.173	59.1	40.9	0.262 (0.053)	0.262	32.2	67.8
Auto	0.070 (0.049)	0.074	62.7	37.3	0.295 (0.047)	0.096	6.1	93.9	0.473 (0.042)	0.040	0.0	100
LS	0.072 (0.048)	0.068	62.9	37.1	0.276 (0.043)	0.072	0.7	99.3	0.474 (0.035)	0.051	0.0	100

Abbreviations: IV, instrumental variable; Est. estimate; SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted; Rr, robust regression; PW, penalized weights; LTS, least trimmed squares; LS, Lasso selection; Auto, automated.

[a]The following two sets of genetic variants were selected from the LTS estimator and included in the IVW model: 1) the h=50 variants used to estimate the initial LTS estimate $\hat{\theta}_{LTS,h}$; and 2) the variants selected from the automated approach based on the heterogeneity stopping rule.

Table 3.11 Results from the simulation study when 100 genetic variants were simulated for 1,000 datasets. Mean (standard error), standard deviation, coverage (%), and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (Rr); 3) penalized weights (PW); 4) robust regression and penalized weights (Rr and PW); 5) the h genetic variants selected by the least trimmed squared (LTS) estimator; and 6) the genetic variants from the Lasso selection (LS) method with the heterogeneity stopping rule for Scenarios 2–4 with a positive causal effect ($\theta = 0.3$) by the number of invalid instrumental variables.

	5 invalid IV				**15 invalid IVs**				**30 invalid IVs**			
	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.
Scenario 2. Balanced pleiotropy, InSIDE satisfied												
IVW	0.227 (0.079)	0.076	17.8	82.2	0.229 (0.108)	0.113	45.5	54.5	0.228 (0.141)	0.136	64.2	35.8
Rr	0.227 (0.065)	0.062	6.1	93.9	0.233 (0.080)	0.082	18.6	81.4	0.230 (0.126)	0.119	54.3	45.7
PW	0.230 (0.061)	0.062	3.9	96.1	0.241 (0.064)	0.079	8.2	91.8	0.241 (0.071)	0.100	15.7	84.3
Rr and PW	0.229 (0.064)	0.063	4.8	95.2	0.237 (0.072)	0.077	11.1	88.9	0.236 (0.088)	0.097	25.5	74.5
LTS[a]												
h	0.235 (0.081)	0.157	30.5	69.5	0.234 (0.082)	0.158	30.9	69.1	0.238 (0.082)	0.145	27.9	72.1
Auto	0.227 (0.062)	0.062	4.9	95.1	0.234 (0.069)	0.084	12.8	87.2	0.232 (0.088)	0.125	31.1	68.9
LS	0.227 (0.061)	0.064	4.8	95.2	0.232 (0.065)	0.079	9.2	90.8	0.230 (0.072)	0.095	17.8	82.2
Scenario 3. Directional pleiotropy, InSIDE satisfied												
IVW	0.323 (0.079)	0.068	1.1	98.9	0.514 (0.107)	0.081	0	100	0.804 (0.133)	0.098	0.0	100
Rr	0.251 (0.066)	0.066	2.9	97.1	0.342 (0.093)	0.081	1.3	98.7	0.654 (0.162)	0.107	0.2	99.8
PW	0.251 (0.061)	0.066	2.1	97.9	0.308 (0.065)	0.080	0.8	99.2	0.490 (0.076)	0.121	0.0	100
Rr and PW	0.246 (0.065)	0.067	4.1	95.9	0.291 (0.074)	0.080	3.2	96.8	0.442 (0.102)	0.123	1.6	98.3
LTS[a]												
h	0.236 (0.081)	0.156	31.2	68.8	0.244 (0.082)	0.156	30.4	69.6	0.253 (0.083)	0.157	27.1	72.9
Auto	0.248 (0.062)	0.065	2.4	97.6	0.299 (0.070)	0.082	1.6	98.4	0.509 (0.091)	0.15	0.3	99.7
LS	0.247 (0.061)	0.066	2.5	97.5	0.297 (0.065)	0.081	1.3	98.7	0.463 (0.074)	0.115	0.0	100
Scenario 4. Directional pleiotropy, InSIDE violated												
IVW	0.411 (0.061)	0.060	0.0	100	0.609 (0.058)	0.054	0.0	100	0.747 (0.051)	0.045	0.0	100
Rr	0.327 (0.095)	0.076	4.5	95.5	0.575 (0.093)	0.065	0.6	99.4	0.746 (0.057)	0.047	0.0	100
PW	0.307 (0.058)	0.080	0.9	99.1	0.479 (0.053)	0.093	0.1	99.9	0.689 (0.046)	0.066	0.0	100
Rr and PW	0.298 (0.072)	0.080	1.7	98.3	0.478 (0.073)	0.098	0.2	99.6	0.697 (0.051)	0.066	0.0	100
LTS[a]												
h	0.239 (0.080)	0.176	31.9	68.1	0.313 (0.076)	0.220	24.7	75.3	0.583 (0.059)	0.274	8.6	91.4
Auto	0.319 (0.058)	0.087	1.1	98.9	0.569 (0.055)	0.088	0.0	100	0.745 (0.049)	0.049	0.0	100
LS	0.314 (0.057)	0.08	0.3	99.7	0.544 (0.050)	0.081	0.0	100	0.742 (0.041)	0.061	0.0	100

Abbreviations: IV, instrumental variable; Est. estimate; SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted; Rr, robust regression; PW, penalized weights; LTS, least trimmed squares; LS, Lasso selection; Auto, automated.

[a]The following three sets of genetic variants were selected from the LTS estimator and included in the IVW model: 1) the h=50 variants used to estimate the initial LTS estimate $\hat{\theta}_{LTS,h}$; and 2) the variants selected from the automated approach based on the heterogeneity stopping rule.

3.5.5 Binary outcome

The data generating model in Section 3.5.1 considered a continuous risk factor X and continuous outcome Y. As noted in Section 2.3.3, the IVW method provides an approximate measure of the causal odds ratio for the effect of a continuous risk factor on a binary outcome. To investigate how well the methods proposed in Section 3.3 estimate the causal odds ratio, we repeated the simulation study outlined in Section 3.5.1 for Scenarios 1–4 for 1 000 simulated datasets with a binary outcome Y.

With the exception of the outcome Y, the same data generating model outlined in Section 3.5.1 was used for the additional simulation. The outcome was generated from the Binomial distribution:

$$Y_i \sim \text{Binomial}(1, \text{expit}(\theta_0 + \sum_{j=1}^{J} \alpha_j G_{ij} + \theta X_i + U_i)) \,, \tag{3.14}$$

where θ represents the log odds ratio per unit increase in the risk factor X, and expit is the inverse of the logit function:

$$\text{expit}(x) = \frac{\exp(x)}{1 + \exp(x)} \,.$$

The constant term θ_0 in Equation 3.14 was set to -3.75 to ensure the prevalence of the outcome Y ranged between 3% and 8% across the different scenarios. The summary level data for the outcome was obtained by fitting logistic regression models for each genetic variant G_j $(j = 1, \ldots, J)$.

We applied the methods discussed in Section 3.3 (summarised in Table 3.1) to the simulated datasets, including the IVW method with: 1) robust regression; 2) penalized weights; 3) robust regression and penalized weights; 4) the three sets of genetic variants from the LTS selection method as outlined in Section 3.3.4; and 5) the genetic variants from the Lasso selection method with the heterogeneity stopping rule. The IVW (with all J genetic variants), simple (unweighted) median, weighted median, and MR-Egger methods were also considered.

Results

The mean R^2 statistic, F-statistic, I^2 statistic and prevalence of the outcome are contained in Table 3.12 for Scenarios 1–4 for the null and positive causal effect by the number of invalid IVs. As expected, the mean values for the R^2 and F-statistics were very similar to those reported in the original simulation (Table 3.3). There was

little difference in the mean I^2 statistic in Tables 3.3 and 3.12. The prevalence of the outcome ranged between 3.5% and 5% for the null causal effect, and 5.8% and 7.9% for the positive causal effect. The prevalence of the outcome generally increased as the number of invalid IVs increased.

Table 3.12 Mean values of the R^2 (%), F-statistic, I^2 (%) and prevalence of the outcome (%) for Scenarios 1-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instrumental variables (IV) when the simulation study was re-performed with a binary outcome for a 1,000 simulated datasets.

	No invalid IVs				1 invalid IV				3 invalid IVs				6 invalid IVs			
	R^2	F	I^2	P	R^2	F	I^2	P	R^2	F	I^2	P	R^2	F	I^2	P
Null causal effect: $\theta = 0$																
Scen 1	3.0	20.9	39.8	3.5	-	-	-	-	-	-	-	-	-	-	-	-
Scen 2	-	-	-	-	3.0	20.7	38.8	3.5	3.1	21.0	40.1	3.6	3.0	21.0	39.0	3.6
Scen 3	-	-	-	-	3.0	20.8	38.9	3.8	3.0	20.6	39.2	4.2	3.0	20.6	40.1	5.0
Scen 4	-	-	-	-	3.4	23.7	55.7	3.7	4.2	29.2	70.9	4.0	5.4	37.8	77.9	4.6
Positive causal effect: $\theta = 0.3$																
Scen 1	3.0	20.7	40.5	5.8	-	-	-	-	-	-	-	-	-	-	-	-
Scen 2	-	-	-	-	3.0	20.8	39.3	5.8	3.0	20.7	40.3	5.9	3.0	20.8	40.7	5.9
Scen 3	-	-	-	-	3.0	20.8	39.9	6.1	3.0	20.8	39.1	6.8	3.1	21.0	40.0	7.9
Scen 4	-	-	-	-	3.4	23.6	56.7	6.1	4.2	29.1	71.4	6.8	5.3	37.5	77.8	7.8

Abbreviations: IV, instrumental variable; Scen, scenario; F, F-statistic; P, prevalence.

Results from the simulation study with a binary outcome for the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (Rr); 3) penalized weights (PW); 4) robust regression and penalized weights; and 5) the three sets of variants selected by LTS selection are provided in Table 3.13 (Scenario 1 only), Table 3.14 (null causal effect $\theta = 0$), and Table 3.15 (positive causal effect $\theta = 0.3$). Tables 3.13 to 3.15 also contain the results from Lasso selection with the heterogeneity stopping rule, simple (unweighted) median, weighted median and MR-Egger methods, and for each method, information on the: mean estimate; mean standard error of the estimates; standard deviation of the estimates; coverage of the true causal effect of the 95% confidence interval; and power to detect the causal effect at the 5% significance level are provided.

The mean estimates of the null causal effect were slightly biased for all of the methods when there were no invalid IVs (Table 3.13). As seen in the original simulation (Table 3.4), the Type 1 error rates for the null causal effect were close to the nominal level of 5% except for the IVW model with the h variants selected from LTS penalization. The attenuation towards the null with a positive causal effect was more severe for a binary outcome (Table 3.13) compared to the original simulation with a

continuous outcome (Table 3.4). Although the coverage of the positive causal effect was close to 95% for the methods considered, there was a significant drop in power when the outcome was binary rather than continuous. The IVW method only had 25.4% power to detect the causal effect, and the simple median had 13.1% power. This reduction in power is reflected in the increase in the mean standard errors for all of the methods for the null and positive causal effect.

Table 3.13 Results from the simulation study when the outcome was binary and 1,000 datasets were simulated. Mean (standard error), standard deviation, coverage (%), and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression; 3) penalized weights; 4) robust regression and penalized weights; 5) the three sets of variants selected by the least trimmed squared (LTS) estimator; and 6) the genetic variants from the Lasso selection method with the heterogeneity stopping rule for Scenario 1 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect. Results from the simple (unweighted) median, weighted median and MR-Egger methods are also provided.

	Null causal effect ($\theta = 0$)				**Positive causal effect** ($\theta = 0.3$)			
	Estimate (SE)	SD	Cov.	Pow.	Estimate (SE)	SD	Cov.	Pow.
Scenario 1. No pleiotropy, InSIDE automatically satisfied								
IVW	0.018 (0.233)	0.214	96.2	3.8	0.249 (0.184)	0.173	94.8	25.4
Robust regression	0.017 (0.252)	0.223	96.0	4.0	0.245 (0.197)	0.177	94.3	25.3
Penalized weights	0.020 (0.230)	0.221	95.6	4.4	0.248 (0.181)	0.176	94.5	26.5
Robust regression with penalized weights	0.018 (0.245)	0.228	95.5	4.5	0.245 (0.192)	0.179	93.7	26.6
LTS[a]								
Variants from h	0.008 (0.299)	0.439	81.7	18.3	0.242 (0.236)	0.350	81.2	28.8
Variants from $w_{LTS,2}$	0.019 (0.232)	0.243	93.9	6.1	0.248 (0.183)	0.187	93.7	26.8
Automated approach	0.021 (0.231)	0.218	95.8	4.2	0.248 (0.182)	0.176	94.4	25.9
Lasso selection	0.023 (0.231)	0.223	95.3	4.7	0.248 (0.182)	0.177	94.3	26.1
Median								
Simple	0.004 (0.330)	0.279	98.4	1.6	0.246 (0.261)	0.228	97.4	12.4
Weighted	0.025 (0.304)	0.267	97.2	2.8	0.236 (0.241)	0.212	97.5	13.1
MR-Egger	-0.012 (0.837)	0.768	97.2	2.8	0.119 (0.652)	0.649	94.9	4.8

Abbreviations: SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse-variance weighted; LTS, least trimmed squares.
[a]The following three sets of genetic variants were selected from the LTS estimator and included in the IVW model: 1) the h=8 variants used to estimate the initial LTS estimate $\hat{\theta}_{LTS,h}$; 2) the variants with a weight of 1 in $w_{LTS,2}$; and 3) the variants selected from the automated approach based on the heterogeneity stopping rule.

Although the mean estimates for the null and positive causal effects in Scenario 2 (Tables 3.14 and 3.15) were similar to those in Scenario 1, most of the methods reported a larger mean standard error under Scenario 2. As with Scenarios 1 and 2, the mean estimates for the null causal effect under Scenarios 3 and 4 were more biased when the outcome was binary rather continuous, and the mean estimates for a positive causal effect had greater attenuation towards the null.

In terms of the performance of the methods, the overall conclusions drawn from the original simulation study are applicable to the results with a binary outcome. In particular, the IVW method with penalized weights had the most precise estimates, but suffered from inflated Type 1 error rates and poor coverage, and robust regression with penalized weights performed reasonably well when there was 1 or 3 invalid IVs. However, as observed in the original simulation, there was little advantage in applying the robust methods compared to the median estimator.

Table 3.14 Results from the simulation study when the outcome was binary and 1,000 datasets were simulated. Mean (standard error), standard deviation, coverage (%), and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (Rr); 3) penalized weights (PW); 4) robust regression and penalized weights (Rr and PW); 5) the three sets of variants selected by the least trimmed squared (LTS) estimator; and 6) the genetic variants from the Lasso selection (LS) method with the heterogeneity stopping rule for Scenarios 2-4 with a null causal effect ($\theta = 0$) by the number of invalid IVs. Results from the simple median, weighted median and MR-Egger methods are also provided.

	1 invalid IV				**3 invalid IVs**				**6 invalid IVs**			
	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.
Scenario 2. Balanced pleiotropy, InSIDE satisfied												
IVW	0.016 (0.241)	0.218	95.9	4.1	0.014 (0.249)	0.248	94.8	5.2	0.016 (0.273)	0.270	93.6	6.4
Rr	0.013 (0.260)	0.226	95.9	4.1	0.012 (0.265)	0.255	93.9	6.1	0.007 (0.293)	0.278	92.6	7.3
PW	0.013 (0.235)	0.224	95.0	5.0	0.010 (0.239)	0.258	92.8	7.2	-0.003 (0.254)	0.296	89.7	10.3
Rr and PW	0.011 (0.251)	0.230	95.0	5.0	0.010 (0.253)	0.262	92.5	7.5	-0.004 (0.267)	0.299	89.6	10.3
LTS[a]												
h	0.007 (0.302)	0.468	80.2	19.8	0.025 (0.297)	0.462	78.5	21.5	-0.006 (0.298)	0.544	71.1	28.9
$w_{LTS,2}$	0.016 (0.238)	0.248	93.0	7.0	0.017 (0.243)	0.268	91.6	8.4	0.014 (0.262)	0.304	88.8	11.2
Auto	0.015 (0.238)	0.223	95.3	4.7	0.009 (0.244)	0.256	93.6	6.4	0.006 (0.264)	0.290	90.8	9.2
LS	0.015 (0.238)	0.229	94.8	5.2	0.010 (0.242)	0.257	93.2	6.8	0.007 (0.262)	0.294	89.9	10.1
Median												
Simple	-0.002 (0.335)	0.290	97.7	2.3	0.000 (0.337)	0.316	96.0	4.0	-0.023 (0.348)	0.355	94.1	5.9
Weighted	0.023 (0.310)	0.274	96.6	3.4	0.032 (0.310)	0.300	95.9	4.1	0.019 (0.318)	0.339	92.7	7.3
MR-Egger	-0.003 (0.872)	0.826	96.4	3.6	0.020 (0.896)	0.850	95.6	4.4	-0.007 (0.986)	1.006	93.8	6.2
Scenario 3. Directional pleiotropy, InSIDE satisfied												
IVW	0.073 (0.232)	0.217	96.0	4.0	0.204 (0.231)	0.209	87.3	12.7	0.379 (0.231)	0.202	64.1	35.9
Rr	0.061 (0.248)	0.225	94.9	5.0	0.178 (0.250)	0.222	88.8	11.2	0.366 (0.248)	0.210	69.6	30.3
PW	0.060 (0.226)	0.223	95.4	4.6	0.182 (0.222)	0.221	86.4	13.6	0.364 (0.217)	0.218	61.5	38.5
Rr and PW	0.057 (0.241)	0.228	94.6	5.3	0.171 (0.239)	0.228	87.6	12.4	0.358 (0.230)	0.223	66.0	34.0
LTS[a]												
h	0.060 (0.291)	0.441	79.5	20.5	0.108 (0.277)	0.443	77.7	22.3	0.306 (0.254)	0.444	64.4	35.6
$w_{LTS,2}$	0.061 (0.229)	0.242	93.4	6.6	0.177 (0.225)	0.240	85.5	14.5	0.372 (0.224)	0.227	61.9	38.1
Auto	0.063 (0.229)	0.223	95.4	4.6	0.189 (0.226)	0.221	87.0	13.0	0.370 (0.225)	0.211	63.3	36.7
LS	0.060 (0.228)	0.225	95.4	4.6	0.188 (0.225)	0.225	86.5	13.5	0.366 (0.222)	0.218	63.1	36.9
Median												
Simple	0.055 (0.323)	0.283	97.9	2.1	0.161 (0.313)	0.280	93.3	6.7	0.362 (0.298)	0.272	80.2	19.8
Weighted	0.063 (0.299)	0.270	97.8	2.2	0.157 (0.288)	0.266	93.0	7.0	0.333 (0.273)	0.262	77.9	22.1
MR-Egger	-0.023 (0.835)	0.859	94.0	6.0	0.001 (0.828)	0.826	94.6	5.4	0.005 (0.817)	0.861	93.5	6.5
Scenario 4. Directional pleiotropy, InSIDE violated												
IVW	0.079 (0.215)	0.205	94.8	5.2	0.182 (0.188)	0.170	85.2	14.8	0.279 (0.158)	0.143	56.7	43.3
Rr	0.073 (0.243)	0.220	93.4	6.5	0.176 (0.213)	0.178	84.1	15.8	0.279 (0.171)	0.151	60.5	39.5
PW	0.068 (0.212)	0.217	93.6	6.4	0.168 (0.185)	0.183	84.9	15.1	0.270 (0.154)	0.150	57.5	42.5
Rr and PW	0.068 (0.233)	0.226	92.4	7.5	0.168 (0.206)	0.187	83.6	16.3	0.273 (0.163)	0.155	59.5	40.5
LTS[a]												
h	0.062 (0.277)	0.426	78.6	21.4	0.122 (0.242)	0.373	75.5	24.5	0.257 (0.197)	0.304	59.7	40.3
$w_{LTS,2}$	0.070 (0.215)	0.238	91.5	8.5	0.168 (0.188)	0.197	84.1	15.9	0.276 (0.156)	0.164	56.7	43.3
Auto	0.072 (0.214)	0.215	94.2	5.8	0.175 (0.186)	0.179	84.7	15.3	0.278 (0.156)	0.149	56.1	43.9
LS	0.069 (0.214)	0.218	93.8	6.2	0.175 (0.186)	0.180	84.1	15.9	0.277 (0.156)	0.151	55.9	44.1
Median												
Simple	0.036 (0.309)	0.282	97.0	3.0	0.115 (0.274)	0.239	94.5	5.5	0.219 (0.226)	0.200	85.5	14.5
Weighted	0.121 (0.282)	0.266	94.8	5.2	0.234 (0.243)	0.217	86.2	13.8	0.328 (0.200)	0.181	64.0	36.0
MR-Egger	0.292 (0.697)	0.680	93.2	6.8	0.504 (0.548)	0.524	85.8	14.2	0.619 (0.455)	0.435	71.5	28.5

Abbreviations: IV, instrumental variable; Est. estimate; SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted; Rr, robust regression; PW, penalized weights; LTS, least trimmed squares; LS, Lasso selection; Auto, automated.

[a]The following three sets of genetic variants were selected from the LTS estimator and included in the IVW model: 1) the h=8 variants used to estimate the initial LTS estimate $\hat{\theta}_{LTS,h}$; 2) the variants with a weight of 1 in $w_{LTS,2}$; and 3) the variants selected from the automated approach based on the heterogeneity stopping rule.

Table 3.15 Results from the simulation study when the outcome was binary and 1,000 datasets were simulated. Mean (standard error), standard deviation, coverage (%), and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (Rr); 3) penalized weights (PW); 4) robust regression and penalized weights (Rr and PW); 5) the three sets of variants selected by the least trimmed squared (LTS) estimator; and 6) the genetic variants from the Lasso selection (LS) method with the heterogeneity stopping rule for Scenarios 2-4 with a positive causal effect ($\theta = 0.3$) by the number of invalid IVs. Results from the simple median, weighted median and MR-Egger methods are also provided.

	1 invalid IV				**3 invalid IVs**				**6 invalid IVs**			
	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.
Scenario 2. Balanced pleiotropy, InSIDE satisfied												
IVW	0.263 (0.190)	0.190	93.8	27.8	0.268 (0.207)	0.207	93.6	28.2	0.274 (0.226)	0.220	94.5	25.7
Rr	0.260 (0.204)	0.196	92.3	27.4	0.266 (0.220)	0.207	93.5	26.9	0.268 (0.237)	0.224	93.9	24.3
PW	0.260 (0.185)	0.199	92.4	29.3	0.262 (0.195)	0.212	92.4	30.9	0.267 (0.205)	0.237	89.6	31.4
Rr and PW	0.259 (0.196)	0.202	91.8	29.3	0.263 (0.206)	0.213	92.4	30.1	0.265 (0.215)	0.238	90.5	29.3
LTS[a]												
h	0.264 (0.235)	0.351	81.4	32.3	0.254 (0.235)	0.376	78.0	34.7	0.271 (0.234)	0.418	72.9	35.5
$w_{LTS,2}$	0.263 (0.187)	0.211	90.6	31.4	0.272 (0.199)	0.216	92.2	30.8	0.275 (0.214)	0.238	90.6	30.7
Auto	0.261 (0.187)	0.198	92.9	28.7	0.264 (0.200)	0.211	92.0	29.2	0.270 (0.215)	0.235	91.3	29.3
LS	0.259 (0.187)	0.197	92.7	28.5	0.263 (0.198)	0.215	91.0	29.7	0.270 (0.212)	0.242	90.3	30.8
Median												
Simple	0.264 (0.264)	0.236	97.6	15.5	0.261 (0.270)	0.244	96.4	13.1	0.270 (0.278)	0.280	95.5	17.5
Weighted	0.255 (0.243)	0.230	96.3	17.9	0.264 (0.249)	0.239	95.6	18.3	0.264 (0.254)	0.263	95.0	19.3
MR-Egger	0.104 (0.686)	0.661	93.8	4.5	0.133 (0.736)	0.733	93.6	5.2	0.100 (0.809)	0.771	94.5	5.0
Scenario 3. Directional pleiotropy, InSIDE satisfied												
IVW	0.315 (0.186)	0.164	96.7	38.4	0.421 (0.190)	0.170	93.0	61.7	0.590 (0.191)	0.163	69.4	89.6
Rr	0.300 (0.198)	0.172	96.4	35.3	0.403 (0.202)	0.178	93.2	53.6	0.578 (0.206)	0.171	72.8	83.5
PW	0.306 (0.179)	0.172	95.7	40.1	0.401 (0.179)	0.180	91.6	62.0	0.577 (0.177)	0.180	65.7	90.0
Rr and PW	0.299 (0.191)	0.176	95.3	37.0	0.395 (0.189)	0.184	91.7	56.6	0.571 (0.188)	0.184	67.6	85.8
LTS[a]												
h	0.293 (0.229)	0.347	79.2	34.3	0.364 (0.218)	0.360	77.0	42.3	0.526 (0.203)	0.385	64.0	60.4
$w_{LTS,2}$	0.298 (0.182)	0.183	94.2	38.0	0.400 (0.183)	0.193	90.7	58.6	0.582 (0.185)	0.184	67.1	88.8
Auto	0.304 (0.182)	0.173	95.7	38.7	0.406 (0.184)	0.180	92.0	60.4	0.582 (0.185)	0.176	68.2	89.1
LS	0.303 (0.181)	0.174	95.4	38.5	0.403 (0.182)	0.183	91.8	60.1	0.579 (0.183)	0.185	67.1	88.7
Median												
Simple	0.312 (0.257)	0.222	98.4	19.6	0.406 (0.252)	0.223	95.5	33.2	0.593 (0.245)	0.228	79.3	69.6
Weighted	0.288 (0.237)	0.213	97.4	20.2	0.381 (0.231)	0.217	96.1	37.8	0.547 (0.222)	0.222	80.3	68.6
MR-Egger	0.082 (0.664)	0.616	93.6	3.4	0.107 (0.684)	0.705	93.4	6.9	0.141 (0.680)	0.686	92.6	7.1
Scenario 4. Directional pleiotropy, InSIDE violated												
IVW	0.320 (0.171)	0.160	95.6	47.0	0.421 (0.150)	0.144	88.9	79.4	0.509 (0.125)	0.116	61.9	98.6
Rr	0.312 (0.195)	0.171	94.6	40.6	0.415 (0.173)	0.154	88.2	70.7	0.507 (0.134)	0.123	63.6	94.4
PW	0.309 (0.168)	0.172	94.2	46.6	0.407 (0.146)	0.159	87.0	78.0	0.500 (0.121)	0.123	61.8	97.3
Rr and PW	0.307 (0.185)	0.176	93.9	42.0	0.406 (0.163)	0.164	86.5	72.6	0.501 (0.127)	0.127	61.8	95.3
LTS[a]												
h	0.289 (0.218)	0.337	79.3	37.4	0.376 (0.189)	0.314	73.4	52.9	0.479 (0.152)	0.251	61.7	76.6
$w_{LTS,2}$	0.307 (0.171)	0.187	92.4	45.2	0.405 (0.148)	0.174	84.8	75.3	0.504 (0.123)	0.135	59.9	96.3
Auto	0.312 (0.169)	0.170	94.4	46.6	0.415 (0.148)	0.157	86.8	78.8	0.508 (0.123)	0.122	60.6	98.0
LS	0.311 (0.169)	0.173	93.9	46.7	0.415 (0.148)	0.157	86.9	78.8	0.508 (0.122)	0.124	59.7	98.1
Median												
Simple	0.294 (0.246)	0.214	97.1	19.3	0.368 (0.217)	0.201	95.5	39.0	0.459 (0.179)	0.162	87.5	75.3
Weighted	0.343 (0.224)	0.201	96.8	31.0	0.458 (0.194)	0.187	87.6	66.8	0.537 (0.159)	0.148	69.9	93.0
MR-Egger	0.406 (0.550)	0.526	94.9	13.1	0.680 (0.432)	0.420	84.8	36.4	0.769 (0.360)	0.346	75.6	57.6

Abbreviations: IV, instrumental variable; Est. estimate; SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted; Rr, robust regression; PW, penalized weights; LTS, least trimmed squares; LS, Lasso selection; Auto, automated.

[a]The following three sets of genetic variants were selected from the LTS estimator and included in the IVW model: 1) the h=8 variants used to estimate the initial LTS estimate $\hat{\theta}_{LTS,h}$; 2) the variants with a weight of 1 in $w_{LTS,2}$; and 3) the variants selected from the automated approach based on the heterogeneity stopping rule.

3.5.6 Summary

In this Section, we have performed an extensive simulation study to compare the performance of the robust methods outlined in Section 3.3 to the IVW, simple median, unweighted median and MR-Egger methods. The study highlighted the sensitivity of the IVW model to violations in the IV assumptions. Due to violations in the NOME assumption, it was not possible to compare the performance of the robust methods to the MR-Egger model. Although the robust methods introduced in this Chapter did not perform significantly better than the median estimator, there appears to be some merit in applying the methods in certain scenarios. For instance, the estimates from robust regression with penalized weights were more precise than the median estimator when there were few invalid instruments. There were no noticeable improvements in the robust methods when the number of genetic variants was increased.

3.6 Discussion

In this Chapter, we have introduced four robust methods for Mendelian randomization with summarized level data that downweight the influence of heterogeneous causal estimates (Section 3.3). The applied examples in Section 3.4 illustrate the importance of using a variety of methods in a Mendelian randomization analysis. The results from the robust methods support a null causal effect of BMI on schizophrenia risk. With the exception of LTS selection with the h variants and Lasso selection with the cross validation stopping rule, the robust methods did not support the positive causal effects estimated from the IVW and MR-Egger methods for LDL-C and AD risk.

In Section 3.5 we performed a simulation study to compare the robust methods to the IVW, simple median, weighted median, and MR-Egger method. The simulation study highlighted the sensitivity of the IVW method to violations in the IV assumptions, and the requirement for robust methods to be considered in the sensitivity analysis of a Mendelian randomization study. The simulations also demonstrated the impact of violating the NOME assumption on the estimates from the MR-Egger method. The SIMEX method [49] described in Section 2.6.2 could have been used to adjust for the effect of violating the NOME assumption on the MR-Egger method. As outlined in Section 2.6.2, the SIMEX method requires the genetic assoications with the risk factor to be simulated multiple times, with the MR-Egger method re-fitted to each simulated dataset. It was decided that this would be too computationally expensive for the simulation study in Section 3.5, and no adjustment was made for the violation of the NOME assumption.

The IVW model with the h variants from the LTS selection method, followed by robust regression with penalized weights, consistently produced the least biased estimates in the simulation study. Although the power and bias of these two methods was significantly better than the IVW method when the IV assumptions were violated, they suffered from poor coverage and increased type I error rates, particularly when there was a high proportion of invalid instruments. When there was only one invalid instrument, robust regression with penalized weights produced more precise estimates than the median estimator. However, when the number of invalid instruments were increased there was little advantage of using robust regression with penalized weights compared to the median estimator.

3.6.1 Comparison with the arXiv paper

As highlighted in the acknowledgments and Section 1.7, this Chapter is based on work originally carried out by Stephen Burgess and other collaborators (Jack Bowden, Frank Dudbridge and Simon Thompson). A copy of a draft manuscript uploaded to arXiv by Stephen Burgess and colleagues on this work has been provided in Appendix A. For clarity, we now highlight the differences between the work presented in this Chapter and the work presented in Appendix A.

Stephen Burgess developed the methods in Sections 3.3.1 to 3.3.3. Although these methods from Appendix A have not been changed, 'penalization of weights' and 'L1 penalization' have been renamed to 'penalized weights' and 'Lasso selection' for the dissertation. The R code for the heterogeneity stopping rule for Lasso selection written by Stephen Burgess was used for the Chapter. Jessica Rees suggested the idea of using the least trimmed squares estimators as a means of selecting genetic variants as IVs, and developed this method independently.

For Section 3.4, the BMI and schizophrenia example is contained in the arXiv paper, and was reanalysed and rewritten by Jessica Rees for the dissertation. The applied example of LDL-C and AD was not considered in the arXiv paper. R code written by Stephen Burgess for producing the Figures 3.2 and 3.3 was used for the Chapter.

Although the simulation study in this Chapter was based on the simulation study in Appendix A, it was been rewritten and extended for the dissertation. Apart from the heterogeneity stopping rule, the R code for the simulation study was written by Jessica Rees. The method used to determine whether the genetic variant was a valid or invalid IV in the simulation was changed, and the parameters used in the data generating model were altered. For the dissertation, Jessica Rees proposed and implemented the idea of presenting the frequency all invalid instruments were either penalized or not selected for the IVW method (Table 3.8). Sections 3.5.4 and 3.5.5 are new additions to the simulation study.

3.6.2 Interpretation of heterogeneity among the causal ratio estimates

Throughout this Chapter, we have assumed that heterogeneity of the causal ratio estimates is indicative of violations in the IV assumptions, particularly the presence of pleiotropic effects. As highlighted in Section 2.5.3, heterogeneity among the causal ratio estimates may arise for a number of reasons [22]. For example, there may be multiple mechanisms of intervention on a complex risk factor, each of which has an

associated causal effect. For a two-sample Mendelian randomization analysis, there may be heterogeneity among the causal ratio estimates due to substantial differences in the study populations used to estimate the genetic associations with the risk factor and outcome. The robust methods considered in this Chapter penalize genetic variants with heterogeneous causal ratio estimates regardless of how this heterogeneity has materialised. As such, these methods should only be employed if it is suspected that the IV assumptions have been violated, and other possible reasons for heterogeneity among the causal ratio estimates explored.

3.6.3 Issues with penalizing genetic variants

The simulation study has highlighted some of the disadvantages of excluding or downweighting genetic variants from Mendelian randomization analyses. Excluding genetic variants from the analysis will generally reduce the standard error of the estimate, resulting in poor coverage of the true causal effect and increased Type 1 error rates, as seen for the LTS and Lasso selection methods. Outlying or heterogeneous causal ratio estimates may be valid IVs, and by excluding them from the analysis we may introduce bias. On balance, it may be more appropriate to consider methods that reduce the contribution that heterogeneous ratio estimates have on the causal estimate, such as the median estimator or robust regression, rather than excluding them from the analysis.

3.6.4 Implication for Mendelian randomization studies

The purpose of this Chapter was not to promote one robust method for Mendelian randomization over another, but to emphasize the need for multiple sensitivity analyses that make different sets of assumptions. Although we acknowledge that none of the proposed methods performed significantly better than the median estimator, the extensions proposed in this Chapter should provide additional confidence in the findings from a conventional Mendelian randomization analysis, particularly when the causal estimates are consistent. Genetic variants that are downweighted or excluded from the analysis by the robust methods should be examined for pleiotropy to determine whether they should be removed from the dataset.

The methods introduced in this Chapter, particularly robust regression with penalized weights, may be more suited to certain scenarios than the median estimator. In the applied example for LDL-C and AD risk, there were two variants that appeared to be clear outliers. The median estimator and robust regression with penalized weights both

suggested that there was a null causal effect of LDL-C on AD risk, but the estimates from the median estimator were less precise. This observation of robust regression producing more precise estimates was also observed in the simulation study when there was one invalid IV. Robust regression with penalized weights may be a useful addition to sensitivity analyses in Mendelian randomization when there are a small proportion of variants with heterogeneous causal estimates.

Based on the work presented in this Chapter, there appears to be little merit in using methods from the robust statistics literature in the sensitivity analysis of a Mendelian randomization study. However, we only considered a limited number of methods from the robust statistics literature, and there may be other methods that are more suitable. Since the simulation study with the increased number of IVs was limited by fixing the R^2 value (discussed in the section below), and the high breakdown points for robust regression and the LTS estimator are based on asymptotic theory, it is difficult to appreciate how these methods may perform for studies with a large number of genetic variants. We did not explore the effect of changing the objective function and/or the value of the tuning parameter for robust regression. The default values used for the tuning parameters in the robust statistics literature for robust regression (used throughout this dissertation) may not be optimal for Mendelian randomization. Based on these observations, and the limitations highlighted in the section below, we decided not to use the robust methods introduced in this Chapter on the main applied example of investigating the effect of adiposity and body composition on asthma (Chapter 6).

3.6.5 Limitations

We found that the Lasso selection method may be more appropriate in an applied setting, where the estimates can be reported over a range of values of the tuning parameter. The practicality of applying Lasso selection to the simulation study was more restrictive, and required an automated approach to selecting the tuning parameter.

Whilst we appreciate the limitation of only considering methods with uncorrelated genetic variants, we argue that robust methods should be used when the IV assumptions are in doubt, and therefore using one genetic variant from each gene region is a sensible approach in an applied Mendelian randomization analysis.

The violation of the NOME assumption limited the utility of the simulation study as the estimates from MR-Egger could not be compared to the robust methods. Given that MR-Egger is frequently used as part of a sensitivity analysis in Mendelian randomization studies, this could be viewed as a major weakness of the simulation study.

The main simulation study was also limited by the number of genetic variants considered in the data generating model. Since GWASs are now being performed on large study populations, and estimates of genetic associations are publicly available from large consortia, only considering 15 variants in the simulation study may have been conservative. We tried to rectify this limitation by re-performing the simulation study with 100 genetic variants and found that there was significant attenuation towards the null due to weak instrument bias. This attenuation was expected as the additional variants did not explain more variance in the risk factor as the R^2 values were kept constant from the original simulation with 15 variants.

We had thought that the performances of some of the robust methods would have improved by increasing the number of genetic variants as the methods are based on asymptotic theory. We did not find any significant improvements in the methods, and in some cases, the performance of the models worsened with the increased number of genetic variants. However, since the R^2 values were set to the same values for the simulations with 15 and 100 genetic variants, limited conclusions can be drawn from the simulation study with the increased number of variants.

This Chapter was also limited by the methods in Section 3.3 only being considered with respect to a continuous outcome. Since many Mendelian randomization studies have a binary outcome, this could be viewed as a fundamental limitation to the work presented. We did try to reduce this limitation by performing additional simulations when the outcome was binary (Section 3.5.5) to see how well the methods approximated the causal odds ratios. Along with the IVW method and the median estimator, the methods introduced in this Chapter produced slightly biased estimates of the null and positive causal effects under all of the scenarios considered. The approximation of the causal odds ratio was worse under a positive effect, with all of the methods reporting mean estimates attenuated towards the null. This attenuation towards the null for the positive causal effect is a result of the non-collapsibility of the odds ratio. The drop in power to detect the positive causal effect for a binary outcome was severe, although the values used to generate the outcome in the data generating model would have influenced this.

3.6.6 Key points from chapter

- In this Chapter, we introduced four extensions to robust methods for Mendelian randomization with summarized data: 1) robust regression (MM-estimation); 2) penalized weights; 3) Lasso selection; and 4) least trimmed squares selection.

These methods can be used to assess the robustness of findings from Mendelian randomization investigations with multiple genetic variants.

- The methods have been considered in two applied examples: one where there is evidence of over-dispersion in the causal estimates (the causal effect of body mass index on schizophrenia risk), and the other containing outliers (the causal effect of low-density lipoprotein on Alzheimer's disease risk).
- Although the performances of the proposed methods in the simulation study were not significantly better than the robust methods that are already in the literature, the applied example suggested that the methods, particularly robust regression with penalized weights, may be a worthwhile addition to a Mendelian randomization study when there is a small proportion of heterogeneous causal estimates.

Chapter 4

Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy

4.1 Introduction

For some sets of risk factors, including lipid fractions, several risk factors have common genetic predictors. Although such genetic variants are pleiotropic, they can be used to estimate causal effects in a multivariable Mendelian randomization framework [28]. In multivariable Mendelian randomization, the IV assumptions are extended to allow a genetic variant to be associated with multiple risk factors, provided all associated risk factors are included in the analysis. Alternatively, when genetic variants are suspected to violate the IV assumptions through unknown pleiotropic pathways, methods have been developed to estimate consistent causal effects under weaker assumptions. These include the weighted median [52] and MR-Egger methods [29].

The extension of MR-Egger to a multivariable setting has been implemented by Helgadottir *et al.* [114] as part of a sensitivity analysis in their applied work investigating the effect of lipid fractions on coronary heart disease (CHD) risk. In the supplementary material, Helgadottir *et al.* [114] suggest that multiple covariates can be included in the MR-Egger method proposed by Bowden *et al.* [29]. The authors then go on to include

summary level data on high-density lipoprotein cholesterol, low-density lipoprotein cholesterol and triglycerides in the same MR-Egger model, and refer to this model as 'multiple Egger regression'. The supplementary material contains the estimates for each lipid fraction from the 'multiple Egger regression' model.

The model we propose in this Chapter for extending MR-Egger to the multivariable setting is essentially the same as the 'multiple Egger regression' model considered by Helgadottir *et al.* [114] in their applied investigation. Since Helgadottir *et al.* [114] only considered the extension of MR-Egger within the context of an applied investigation, there remains several methodological issues relating to the implementation of the method, and the assumptions required. In particular, Helgadottir *et al.* [114] did not address the issues of residual pleiotropic effects, orientation of the genetic variants, or the assumptions required to obtain consistent causal effects.

In this Chapter, we extend the MR-Egger method to the multivariable setting through theoretical arguments (Section 4.2). In Section 4.3, we provide an example analysis using published summary level data on lipid fractions and CHD risk. We also perform two simulation studies to compare the performance of the methods: one where the risk factors do not have causal effects on each other (Section 4.4); and another where this assumption is relaxed (Section 4.5). Finally, in Section 4.6 we discuss the results from Sections 4.2 to 4.5, and consider the implications of extending the MR-Egger method to the multivariable setting on future research.

4.2 Methods

In this Section, we use theoretical arguments to expand the MR-Egger method to the multivariable setting. Initially, we consider the causal effect of a risk factor X on an outcome Y using J genetic variants G_j $(j = 1, \ldots, J)$ that are assumed to be uncorrelated (not in linkage disequilibrium). Then, we expand to consider multiple risk factors $X_1, X_2, \ldots, X_K$, and from the MR-Egger and multivariable IVW models we outline the regression model for multivariable MR-Egger. We provide the assumptions required for the causal estimates from multivariable MR-Egger to be consistent, and compare the precision of the causal estimates from univariable MR-Egger and multivariable MR-Egger. The advantages of using multivariable MR-Egger over univariable MR-Egger are discussed in detail. Finally, we provide a recommendation on how the reference alleles of the genetic variants should be orientated in multivariable MR-Egger.

We assume that summarized data are available on the associations of each genetic variant with the risk factor and with the outcome: the beta-coefficients ($\hat{\beta}_{X_j}, \hat{\beta}_{Y_j}$) and their standard errors ($\text{se}(\hat{\beta}_{X_j}), \text{se}(\hat{\beta}_{Y_j})$) from univariable regression on each variant G_j in turn. For the multivariable setting, we assume that there is summary level data for the genetic variants on each risk factor, and these genetic variants are associated with at least one of the risk factors, and the risk factors are associated with at least one of the genetic variants. Finally, we assume that the associations of genetic variants with the risk factor and the outcome, and the causal effect of the risk factor on the outcome, are linear and homogeneous across the population. To distinguish between the parameters from the different methods considered, we use the following subscript notation: UI ('univariable inverse variance weighted (IVW)'); UE ('univariable MR-Egger'); MI (multivariable IVW); and ME ('multivariable MR-Egger').

4.2.1 Univariable Mendelian randomization

In a univariable Mendelian randomization analysis, each genetic variant must satisfy the IV assumptions outlined in Section 1.4. Under linearity assumptions, the association between the genetic variant and the outcome can be decomposed into an indirect effect via the risk factor and a direct effect:

$$\beta_{Yj} = \alpha_j + \theta\beta_{X_j}\,,$$

where θ is the causal effect of the risk factor on the outcome. Genetic variant j is pleiotropic if $\alpha_j \neq 0$, and α_j is the direct effect of the genetic variant on the outcome. Figure 4.1 contains a direct effect α_j via an independent pathway, which violates the IV3 assumption.

The causal effect of the risk factor on the outcome can be estimated using a weighted linear regression of the genetic association estimates [45], with the inverse-variance as weights ($\text{se}(\hat{\beta}_{Y_j})^{-2}$) and the intercept set to zero (the IVW method):

$$\hat{\beta}_{Y_j} = \theta_{UI}\hat{\beta}_{X_j} + \epsilon_{UI_j}\,, \quad \epsilon_{UI_j} \sim \mathcal{N}(0, \phi_{UI}^2\,\text{se}(\hat{\beta}_{Y_j})^2)\,, \tag{4.1}$$

where θ_{UI} is the IVW estimate, ϵ_{UI_j} represents the error term, and ϕ_{UI} represents the residual standard error. Under a multiplicative random-effects model (see Section 2.3.1 for more detail), the residual standard error (ϕ_{UI}) can be greater than one, allowing for heterogeneity among the causal estimates. The point estimate from the fixed- and

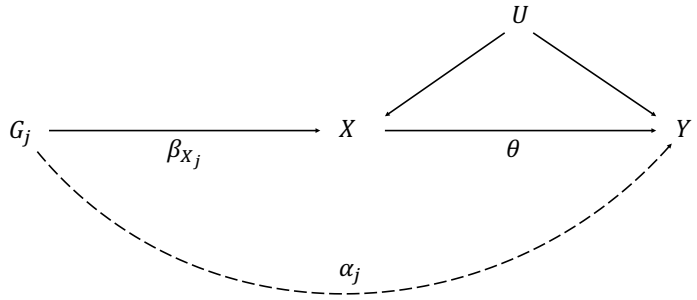


Fig. 4.1 Directed acyclic graph illustrating univariable Mendelian randomization assumptions with potential violation of IV3 by a pleiotropic effect indicated by a dotted line. The genetic effect of G_j on X is β_{X_j}, the direct (pleiotropic) effect of G_j on Y via an independent pathway is α_j (representing the potential violation of the IV3 assumption), and the causal effect of the risk factor X on the outcome Y is θ. U represents the set of unmeasured variables that confound the association between X and Y.

random-effect models will be the same, but the standard error of the causal effect from the multiplicative random-effects model will be larger if there is heterogeneity between the causal estimates. Throughout this Chapter, we apply a multiplicative random-effects model to all the analyses.

The MR-Egger estimate is obtained using the same regression model as Equation 4.1, but allowing the intercept to be estimated [29]:

$$\hat{\beta}_{Y_j} = \theta_{0UE} + \theta_{UE}\hat{\beta}_{X_j} + \epsilon_{UE_j}\,, \quad \epsilon_{UE_j} \sim \mathcal{N}(0, \phi^2_{UE}\,\mathrm{se}(\hat{\beta}_{Y_j})^2)\,,$$

where θ_{UE} is the MR-Egger estimate, ϵ_{UE_j} represents the error term, and ϕ_{UE} represents the residual standard error under the univariable MR-Egger model. If the genetic variants are not pleiotropic, then the intercept term should tend to zero as the sample size increases, and the MR-Egger estimate ($\hat{\theta}_{UE}$) and the IVW estimate ($\hat{\theta}_{UI}$) are both consistent estimates of the causal effect. If the genetic variants are pleiotropic, and the InSIDE assumption holds, then the MR-Egger estimate will be a consistent estimate of θ [29, 64].

Under the InSIDE assumption, the intercept term $\hat{\theta}_{0UE}$ can be interpreted as an estimate of the average direct effect of the genetic variants [52]. If the average direct effect is zero (balanced pleiotropy), and the InSIDE assumption is satisfied, the intercept term should tend to zero as the sample size increases, and the MR-Egger estimate ($\hat{\theta}_{UE}$) and the IVW estimate ($\hat{\theta}_{UI}$) are both consistent estimates of the causal effect. If the intercept term differs from zero, then either the InSIDE assumption is

violated or the average direct effect differs from zero (directional pleiotropy); this is a test of the validity of the IV assumptions (the MR-Egger intercept test).

4.2.2 Multivariable Mendelian randomization

In a multivariable Mendelian randomization analysis, each genetic variant must satisfy the IV assumptions outlined in Section 2.6.3. Now, the association of the genetic variants with the outcome can be decomposed into indirect effects via each of the risk factors and a residual direct effect α'_j. Assuming there are three risk factors and all relationships are linear:

$$\beta_{Y_j} = \alpha'_j + \theta_1\beta_{X_{1j}} + \theta_2\beta_{X_{2j}} + \theta_3\beta_{X_{3j}}\,, \tag{4.2}$$

where θ_k is the causal effect of the risk factor k on the outcome (Figure 4.2). We assume that the risk factors do not have causal effects on each other; we later relax this assumption and allow for causal effects between the risk factors.

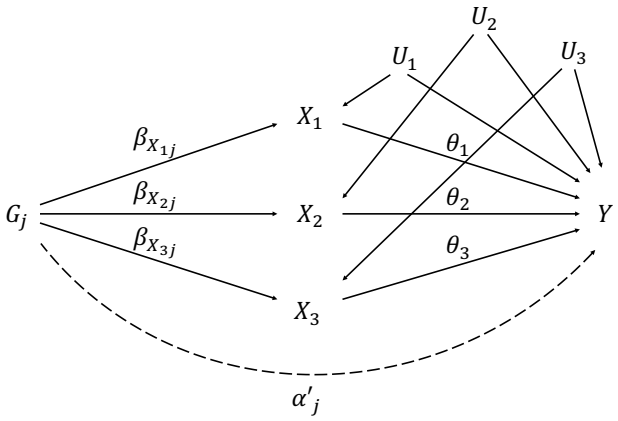


Fig. 4.2 Directed acyclic graph illustrating multivariable Mendelian randomization assumptions for a set of genetic variants G_j, three risk factors X_1, X_2 and X_3, and outcome Y. The genetic effect of G_j on X_k is $\beta_{X_{kj}}$, the direct (pleiotropic) effect of G_j on Y is α'_j, and the causal effect of the risk factor X_k on the outcome Y is θ_k. U_k represents the set of unmeasured variables that confound the associations between X_k and Y.

Causal estimates of the effect of each risk factor on the outcome can be obtained using multivariable weighted linear regression of the genetic association estimates, with the intercept set to zero (multivariable IVW method) [115]:

$$\hat{\beta}_{Y_j} = \theta_{1MI}\hat{\beta}_{X_{1j}} + \theta_{2MI}\hat{\beta}_{X_{2j}} + \theta_{3MI}\hat{\beta}_{X_{3j}} + \epsilon_{MI_j}\,, \quad \epsilon_{MI_j} \sim \mathcal{N}(0, \phi^2_{MI}\,\mathrm{se}(\hat{\beta}_{Y_j})^2)\,,$$

where $\boldsymbol{\theta_{MI}}$ are the multivariable IVW estimates, ϵ_{MI_j} represents the error term, and ϕ_{MI} represents the residual standard error under the multivariable IVW model.

We propose the natural extension to multivariable MR-Egger using the same regression model, but allowing the intercept to be estimated:

$$\hat{\beta}_{Y_j} = \theta_{0ME} + \theta_{1ME}\hat{\beta}_{X_{1j}} + \theta_{2ME}\hat{\beta}_{X_{2j}} + \theta_{3ME}\hat{\beta}_{X_{3j}} + \epsilon_{ME_j}, \quad \epsilon_{ME_j} \sim \mathcal{N}(0, \phi^2_{ME}\,\mathrm{se}(\hat{\beta}_{Y_j})^2)\,.$$

4.2.3 Assumptions for multivariable MR-Egger

We assume that the causal effect of risk factor one (θ_1) is of interest, and provide the assumptions necessary for the multivariable MR-Egger estimate of θ_1 to be consistent. If all of the causal effects are to be interpreted then these assumptions must apply for each risk factor.

If the $\boldsymbol{\beta_{X_1}}$ parameters are independent of the $\boldsymbol{\beta_{X_k}}$ parameters for all $k = 2, 3, \ldots, K$, then the InSIDE assumption for multivariable MR-Egger is satisfied if the direct effects of the genetic variants $\boldsymbol{\alpha'}$ are independent of $\boldsymbol{\beta_{X_1}}$. More formally, we require:

$$\boldsymbol{\beta_{X_1}} \perp\!\!\!\perp \boldsymbol{\alpha'}, \quad \text{if } \boldsymbol{\beta_{X_1}} \perp\!\!\!\perp \boldsymbol{\beta_{X_2}}, \ldots, \boldsymbol{\beta_{X_K}},$$

for the estimate of θ_1 from multivariable MR-Egger to be consistent. If the InSIDE assumption is satisfied, then the weighted covariance of $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\alpha'}$ ($\text{cov}_w(\boldsymbol{\alpha'}, \boldsymbol{\beta_{X_1}})$) will tend to zero as the number of genetic variants J tends to infinity. The estimate of θ_1 from multivariable MR-Egger when the $\boldsymbol{\beta_{X_1}}$ parameters are independent of $\boldsymbol{\beta_{X_k}}$ for all $k = 2, 3, \ldots, K$ is:

$$\hat{\theta}_{1ME} = \frac{\text{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{Y}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X_1}})}{\text{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X_1}})} \xrightarrow{N\to\infty} \frac{\text{cov}_w(\boldsymbol{\beta_Y}, \boldsymbol{\beta_{X_1}})}{\text{var}_w(\boldsymbol{\beta_{X_1}})} = \theta_1 + \frac{\text{cov}_w(\boldsymbol{\alpha'}, \boldsymbol{\beta_{X_1}})}{\text{var}_w(\boldsymbol{\beta_{X_1}})},$$

which is equal to θ_1 if the InSIDE assumption is satisfied, where cov_w and var_w represent the weighted covariance and weighted variance using the inverse-variance weights $\text{se}(\hat{\beta}_{Yj})^{-2}$:

$$\begin{aligned}
\text{cov}_w(\boldsymbol{\alpha'}, \boldsymbol{\beta_{X_1}}) &= \frac{\sum_j (\alpha'_j - \bar{\alpha'}_w)(\beta_{X_1 j} - \bar{\beta}_{X_1 w})\,\text{se}(\hat{\beta}_{Yj})^{-2}}{\sum_j \text{se}(\hat{\beta}_{Yj})^{-2}}, \\
\text{var}_w(\boldsymbol{\beta_{X_1}}) &= \frac{\sum_j (\beta_{X_1 j} - \bar{\beta}_{X_1 w})^2\,\text{se}(\hat{\beta}_{Yj})^{-2}}{\sum_j \text{se}(\hat{\beta}_{Yj})^{-2}}, \\
\bar{\alpha'}_w &= \frac{\sum_j \alpha'_j\,\text{se}(\hat{\beta}_{Yj})^{-2}}{\sum_j \text{se}(\hat{\beta}_{Yj})^{-2}}, \\
\bar{\beta}_{X_1 w} &= \frac{\sum_j \beta_{X_1 j}\,\text{se}(\hat{\beta}_{Yj})^{-2}}{\sum_j \text{se}(\hat{\beta}_{Yj})^{-2}}.
\end{aligned}$$

If the $\boldsymbol{\beta_{X_1}}$ parameters are correlated with at least one of the sets of $\boldsymbol{\beta_{X_k}}$ parameters ($k = 2, 3, \ldots, K$), then the InSIDE assumption is required to hold for $\boldsymbol{\beta_{X_1}}$ and for all

of the $\boldsymbol{\beta_{X_k}}$ parameters that are correlated with $\boldsymbol{\beta_{X_1}}$. More formally, we require:

$$\boldsymbol{\beta_{X_k}} \perp\!\!\!\perp \boldsymbol{\alpha'}, \quad \text{for all } \boldsymbol{\beta_{X_k}} \text{ correlated with } \boldsymbol{\beta_{X_1}} \text{ (including } \boldsymbol{\beta_{X_1}} \text{itself)}.$$

For example, if $k = 2$, and $\boldsymbol{\beta_{X_1}}$ is correlated with $\boldsymbol{\beta_{X_2}}$, we require both of the weighted covariances of $\boldsymbol{\alpha'}$ with $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ to be zero to produce a consistent estimate of θ_1. The estimate of θ_1 from multivariable MR-Egger with two risk factors where $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ are correlated is:

$$\begin{aligned}
\hat{\theta}_{1ME} &= \frac{\mathrm{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{Y}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X_1}})\,\mathrm{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X_2}}) - \mathrm{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{Y}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X_2}})\,\mathrm{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X_1}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X_2}})}{\mathrm{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X_1}})\,\mathrm{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X_2}}) - \mathrm{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X_1}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X_2}})^2} \\
&\xrightarrow{N\to\infty} \frac{\mathrm{cov}_w(\boldsymbol{\beta_Y}, \boldsymbol{\beta_{X_1}})\,\mathrm{var}_w(\boldsymbol{\beta_{X_2}}) - \mathrm{cov}_w(\boldsymbol{\beta_Y}, \boldsymbol{\beta_{X_2}})\,\mathrm{cov}_w(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})}{\mathrm{var}_w(\boldsymbol{\beta_{X_1}})\,\mathrm{var}_w(\boldsymbol{\beta_{X_2}}) - \mathrm{cov}_w(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})^2} \\
&= \theta_1 + \frac{\mathrm{cov}_w(\boldsymbol{\alpha'}, \boldsymbol{\beta_{X_1}})\,\mathrm{var}_w(\boldsymbol{\beta_{X_2}}) - \mathrm{cov}_w(\boldsymbol{\alpha'}, \boldsymbol{\beta_{X_2}})\,\mathrm{cov}_w(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})}{\mathrm{var}_w(\boldsymbol{\beta_{X_1}})\,\mathrm{var}_w(\boldsymbol{\beta_{X_2}}) - \mathrm{cov}_w(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})^2}, \qquad (4.3)
\end{aligned}$$

which is equal to θ_1 if the InSIDE assumption holds with respect to $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$. As more risk factors with correlated sets of association parameters with $\boldsymbol{\beta_{X_1}}$ are included in the multivariable MR-Egger model, additional terms will be added to the bias term in Equation 4.3, and the InSIDE assumption must hold for these additional risk factors to obtain a consistent estimate of θ_1.

4.2.4 Precision of the multivariable MR-Egger estimate

The variance of the multivariable MR-Egger estimate $\hat{\theta}_{1ME}$ will be heavily influenced by the denominator in the bias term of Equation 4.3. The variance of the multivariable MR-Egger estimate $\hat{\theta}_{1ME}$ is given by:

$$\begin{aligned}
\mathrm{var}(\hat{\theta}_{1ME}) &= \frac{\phi^2_{ME}\,\mathrm{var}(\boldsymbol{\beta_{X_2}})}{N(\mathrm{var}(\boldsymbol{\beta_{X_1}})\,\mathrm{var}(\boldsymbol{\beta_{X_2}}) - \mathrm{cov}(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})^2)} \\
&\propto [\mathrm{var}(\boldsymbol{\beta_{X_1}})(1 - \mathrm{cor}(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})^2)]^{-1},
\end{aligned}$$

where ϕ_{ME} is the residual standard error under the multivariable MR-Egger model.

Under a fixed-effect model, the variance of the univariable MR-Egger estimate is proportional to the inverse of $\mathrm{var}(\boldsymbol{\beta_{X_1}})$ [49]. The estimate from the multivariable MR-Egger model $\hat{\theta}_{1ME}$ will be more precise than its univariable counterpart $\hat{\theta}_{1UE}$ if:

$$\frac{1}{\mathrm{var}(\boldsymbol{\beta_{X_1}})} > \frac{1}{\mathrm{var}(\boldsymbol{\beta_{X_1}})(1 - \mathrm{cor}(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})^2)}.$$

From the above inequality, $\hat{\theta}_{1UE}$ will always be more precise than $\hat{\theta}_{1ME}$ when $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ are correlated. Under a multiplicative random-effects model (used throughout this Chapter), the residual standard error is estimated under the univariable MR-Egger model (ϕ_{UE}) and the multivariable MR-Egger model (ϕ_{ME}). For $\hat{\theta}_{1ME}$ to be more precise than $\hat{\theta}_{1UE}$, we require:

$$\frac{\phi^2_{UE}}{\text{var}(\boldsymbol{\beta_{X_1}})} > \frac{\phi^2_{ME}}{\text{var}(\boldsymbol{\beta_{X_1}})(1-\text{cor}(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})^2)} .$$

If $\boldsymbol{\beta_{X_2}}$ explains additional independent variability in the genetic associations with the outcome $\boldsymbol{\beta_Y}$, and $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ are independent, then the estimate from multivariable MR-Egger will be more precise than the estimate from univariable MR-Egger. If $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ are correlated, then the precision of $\hat{\theta}_{1ME}$ will depend upon the strength of the correlation between $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$, and the amount of additional independent variability $\boldsymbol{\beta_{X_2}}$ explains in $\boldsymbol{\beta_Y}$. As the correlation between $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ increases, and $\boldsymbol{\beta_{X_2}}$ explains no additional independent variability in $\boldsymbol{\beta_Y}$, the precision of the multivariable MR-Egger estimate $\hat{\theta}_{1ME}$ will reduce.

4.2.5 Advantages of multivariable MR-Egger and comparison with univariable MR-Egger

The bias for the causal estimate from univariable MR-Egger $\hat{\theta}_{UE}$ depends on the weighted covariance between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta_{X_1}}$, where:

$$\alpha_j = \alpha'_j + \sum_{i=2}^{K} \theta_i \beta_{X_{ij}} . \tag{4.4}$$

The expression in Equation 4.4 follows from the multivariable framework outlined in Equation 4.2, where the direct effect for univariable MR-Egger has been decomposed into the residual direct effect α'_j of multivariable MR-Egger and the indirect effects via each risk factor. The residual direct effect α'_j will be altered with each additional risk factor included in the multivariable MR-Egger model. If these additional risk factors are causally associated with the outcome ($\theta_k \neq 0$), then α'_j will consist of fewer components.

It seems likely that the InSIDE assumption would be easier to satisfy for multivariable MR-Egger than its univariable counterpart as the direct effect for univariable MR-Egger consists of unmeasured and measured pleiotropy. This observation is based on the idea that the magnitude of the direct effect may be smaller for multivariable

MR-Egger compared to its univariable counterpart. However, the remaining direct effect for multivariable MR-Egger may still contain the component that invalidates the InSIDE assumption for univariable MR-Egger, and this would also invalidate the InSIDE assumption for multivariable MR-Egger. Hence, there is no guarantee that the InSIDE assumption will be more easily satisfied under the multivariable setting.

If the $\boldsymbol{\beta_{X_1}}$ parameters are independent of the $\boldsymbol{\beta_{X_k}}$ parameters for all $k = 2, 3, \ldots, K$, then the second term in Equation 4.4 (the measured direct effect) does not contribute to the value of $\text{cov}_w(\boldsymbol{\alpha}, \boldsymbol{\beta_{X_1}})$. Under this scenario, bias for the univariable and multivariable MR-Egger estimates depends on the same covariance term $\text{cov}_w(\boldsymbol{\alpha'}, \boldsymbol{\beta_{X_1}})$. As a consequence, the estimates of the causal effects from univariable MR-Egger $\hat{\theta}_{UE}$ and multivariable MR-Egger $\hat{\theta}_{1ME}$ will be asymptotically the same. In this case, multivariable MR-Egger may improve precision of the causal estimate, but will not affect the asymptotic bias.

When the $\boldsymbol{\beta_{X_1}}$ parameters are correlated with at least one of the sets of $\boldsymbol{\beta_{X_k}}$ parameters for $k = 2, 3, \ldots, K$, the second term in Equation 4.4 now contributes to the value of $\text{cov}_w(\boldsymbol{\alpha}, \boldsymbol{\beta_{X_1}})$. The InSIDE assumption for univariable MR-Egger will therefore be automatically violated as the weighted covariance between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta_{X_1}}$ will not equal zero, resulting in biased causal estimates of θ_1. If the InSIDE assumption holds for multivariable MR-Egger, and $\boldsymbol{\beta_{X_k}}$ are included in the analysis model, then $\hat{\theta}_{1ME}$ will still be a consistent estimate of θ_1. Hence, in this case, multivariable MR-Egger should result in reduced bias compared with univariable MR-Egger.

4.2.6 Orientation of the genetic variants

Genetic associations represent the average change in the risk factor or the outcome per additional copy of the reference allele (e.g. the major or minor allele). There is no biological rationale why associations should be expressed with respect to either the major (wildtype) or the minor (variant) allele. In the univariable and multivariable IVW methods, the estimate is not affected by the choice of orientation, as the intercept is fixed at zero. However, in the univariable and multivariable MR-Egger methods, changing the orientation of the variant affects the intercept term and the causal estimate as the orientation affects the definition of the pleiotropy terms α_j and α'_j. For example, if the genetic variants are orientated with respect to the major alleles, then α_j and α'_j represent the average pleiotropic effect for the univariable and multivariable models with respect to the major alleles. If we were to change the orientation of the genetic variants to the minor alleles, then α_j and α'_j represent the average pleiotropic effects with respect to the minor alleles. Since the orientation of the genetic variants changes

the definition of the pleiotropic effects, the InSIDE assumption will differ with each orientation, and this may also affect the causal estimate.

To ensure that the MR-Egger analysis does not depend on the reported reference alleles, Bowden *et al.* [29] suggested the genetic variants in univariable MR-Egger be orientated so the direction of association with the risk factor is either positive for all variants or negative for all variants. However, this may not be possible for multivariable MR-Egger as the same reference allele must be used for associations with each risk factor and with the outcome. We suggest that the variants should be orientated with respect to their associations with the risk factor of primary interest, although we would recommend a sensitivity analysis considering different orientations if multiple risk factors are of interest. If the genetic variants are all valid instruments, then directional pleiotropy should not be detected with respect to any orientation.

4.2.7 Summary

In this Section, we have expanded MR-Egger to the multivariable setting. We have outlined the assumptions required to obtain consistent causal estimates from multivariable MR-Egger and provided a recommendation on the orientation of the genetic variants. We have also discussed the potential benefits of using multivariable MR-Egger over its univariable counterpart. In the next Section, we will apply the methods discussed to published summary level data to investigate the causal effect of cholesterol on CHD risk.

4.3 Mendelian randomization analysis on the causal effect of high-density lipoprotein cholesterol on coronary heart disease risk

The effects of high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglycerides on the risk of CHD have been investigated by numerous Mendelian randomization studies [116]. For HDL-C, univariable Mendelian randomization suggested a causally protective role against CHD risk, whereas univariable MR-Egger provided no evidence of a causal effect and the test for directional pleiotropy was statistically significant at the 5% level [52]. A null causal effect for HDL-C was also reported from a multivariable Mendelian randomization analysis that included LDL-C and triglycerides using the multivariable IVW method [28], although a small but protective causal effect was estimated in a further multivariable Mendelian randomization analysis using a wider range of 185 genetic variants [117]. In this Section, we investigate the causal effect of HDL-C on CHD risk further using the multivariable MR-Egger method described in Section 4.2.

4.3.1 Methods

We consider the 185 genetic variants having known association with at least one of HDL-C, LDL-C and triglycerides at GWAS significance in 188,578 participants reported by the Global Lipids Genetics Consortium [111]. The point estimates for the associations between these genetic variants and lipids were taken from Do *et al.* [110]. The CARDIoGRAMplusC4D consortium consisting of 60,801 cases and 123,504 controls was used to obtain the estimates of the association between the variants and CHD risk [118]. The IVW and MR-Egger methods were applied to the data under univariable and multivariable frameworks as described in Section 4.2. For the univariable IVW and MR-Egger methods, the models were fitted using two sets of variants: firstly using all 185 variants; and secondly using all variants associated with HDL-C at GWAS significance (p-value $< 5 \times 10^{-8}$). The genetic variants were orientated with respect to the risk increasing allele for HDL-C. These analyses differ from those provided in Burgess *et al.* [117] and Do *et al.* [110] as they use summarized data from different versions of the CARDIoGRAMplusC4D study; here we use associations from the 2015 data release [118].

As a sensitivity analysis, the multivariable MR-Egger method was re-performed with the genetic variants orientated with respect to the risk increasing alleles for LDL-C and triglycerides.

Throughout this Chapter, we have assumed that the associations of the genetic variants with the risk factor and the outcome, and the causal effect of the risk factor on the outcome, are linear and homogeneous. Since the outcome for this applied example is binary, the linearity assumption is violated. The estimates of the causal effects from the analyses will therefore be approximations of the causal odds ratios (Section 2.3.3).

The code used for the analysis was written and performed by Jessica Rees in RStudio version 3.5.3 [86].

4.3.2 Results

The univariable IVW method suggested a significant protective effect of HDL-C for both sets of variants with an approximate causal odds ratio of 0.88 (95% CI: 0.80, 0.97) for all variants (Table 4.1). This estimate attenuated to the null in the univariable MR-Egger method (0.98, 95% CI: 0.87, 1.11) with evidence of directional pleiotropy (p-value=0.004). The approximate causal odds ratios from multivariable IVW (0.96, 95% CI: 0.89, 1.05) and multivariable MR-Egger (1.04, 95% CI: 0.94, 1.14) had opposite directions of association, with both analyses indicating that HDL-C is not causally associated with CHD risk. The significant result for directional pleiotropy in the multivariable MR-Egger method suggests that LDL-C and triglycerides do not fully explain the direct effects of the genetic variants on the outcome, suggesting that there is still residual pleiotropy via other unmeasured risk factors and/or the InSIDE assumption is violated.

The results from the analysis correspond with previous findings that HDL-C has a causal effect on CHD when considered in isolation and without accounting for pleiotropic effects. Univariable MR-Egger, multivariable IVW and multivariable MR-Egger are all consistent with previous findings that HDL-C does not have a causal effect on CHD. Unlike previous studies, we have been able to consider the possibility of there being residual pleiotropy in the multivariable model with HDL-C, LDL-c and triglycerides. Although the significant result for directional pleiotropy in multivariable MR-Egger is of interest, it is not possible to determine whether this is due to residual pleiotropy and/or the InSIDE assumption being violated.

Table 4.1 Approximate log causal odds ratios (95% confidence intervals) for coronary heart disease per standard deviation increase in HDL-C. Estimates of the intercept are given in univariable and multivariable MR-Egger.

	Approximate causal estimate			**MR-Egger intercept test**		
	$\hat{\theta}_{\text{HDL-C}}$ (CI)	se($\hat{\theta}_{\text{HDL-C}}$)	p-value	$\hat{\theta}_{0\text{E}}$	se($\hat{\theta}_{0\text{E}}$)	p-value
UV IVW						
All variants	-0.130 (-0.227, -0.033)	0.049	0.009	-	-	-
Reduced set[a]	-0.114 (-0.211, -0.017)	0.049	0.022	-	-	-
UV MR-Egger						
All variants	-0.016 (-0.138, 0.106)	0.062	0.800	-0.007	0.002	0.004
Reduced set[a]	0.067 (-0.070, 0.204)	0.069	0.332	-0.012	0.004	0.001
MV IVW	-0.039 (-0.123, 0.045)	0.042	0.359	-	-	-
MV MR-Egger	0.036 (-0.063, 0.134)	0.050	0.477	-0.005	0.002	0.008

Abbreviations: MR, Mendelian randomization; UV, univariable; MV, multivariable; HDL-C, high-density lipoprotein cholesterol; IVW, inverse-variance weighted; CI, confidence interval; SE, standard error.
[a]95 variants associated with HDL-C at a genome-wide level of significance (p-value$< 5 \times 10^{-8}$).

Varying the orientation of the genetic variants

The approximate causal estimates for HDL-C, LDL-C, and triglycerides from multivariable MR-Egger when the variants were orientated with respect to HDL-C, LDL-C or triglycerides are presented in Table 4.2. Estimates of the MR-Egger intercept are also provided for the three models. To allow for comparisons between the multivariable methods, the approximate causal estimates from multivariable IVW are included in Table 4.2. The approximate causal estimates in bold follow the recommendation outlined in Section 4.2.6 that the genetic variants should be orientated with respect to the risk factor-increasing allele for the risk factor of interest.

All of the approximate causal odds ratios for HDL-C from the multivariable MR-Egger models indicated that HDL-C is not causally associated with CHD risk. Significant adverse effects of LDL-C on CHD risk were reported from the multivariable IVW (1.45, 95% CI: 1.34, 1.58) and multivariable MR-Egger (1.52, 95% CI: 1.37, 1.69) methods. Orientating the variants with respect to the risk increasing alleles for HDL-C and triglycerides had little impact on the approximate causal estimates for LDL-C from multivariable MR-Egger. The multivariable IVW method suggested a significant adverse effect of triglycerides on CHD risk with an approximate causal odds ratio of 1.19 (95% CI: 1.07, 1.33), this estimate was attenuated to the null in the multivariable MR-Egger method (1.09, 95% CI: 0.96, 1.23). The approximate causal odds ratios for

triglycerides remained significant, however, when the variants were orientated with respect to HDL-C and LDL-C in the multivariable MR-Egger models.

Since the orientation of the genetic variants affects the interpretation of the direct effect, and the definition of the InSIDE assumption, the MR-Egger intercept will vary between different orientations. In this example, the MR-Egger intercept differed from zero when the variants were orientated with respect to HDL-C and triglycerides, yet there was no evidence of directional pleiotropy or the InSIDE assumption being violated when the variants were orientated with respect to LDL-C.

Table 4.2 Approximate causal log odds ratios (95% confidence intervals) for coronary heart disease per standard deviation increase in HDL-C, LDL-C, and triglycerides from multivariable IVW and multivariable MR-Egger. Estimates from multivariable MR-Egger are presented from three models where the reference allele is the risk increasing allele for HDL-C, LDL-C or triglycerides. Estimates of the intercept are given for multivariable MR-Egger.

	Approximate causal estimates			**MR-Egger intercept**
	$\hat{\boldsymbol{\theta}}_{\text{HDL-C}}$	$\hat{\boldsymbol{\theta}}_{\text{LDL-C}}$	$\hat{\boldsymbol{\theta}}_{\text{TG}}$	$\hat{\boldsymbol{\theta}}_{0E}$
MV IVW	-0.039 (-0.123, 0.045)	0.375 (0.292, 0.457)	0.173 (0.063, 0.283)	-
MV MR-Egger[a]				
HDL-C	**0.036 (-0.063, 0.134)**	0.378 (0.297, 0.458)	0.136 (0.024, 0.247)	-0.005 (-0.009, -0.001)
LDL-C	-0.034 (-0.118, 0.049)	**0.420 (0.318, 0.522)**	0.194 (0.081, 0.308)	-0.003 (-0.007, 0.001)
TG	-0.018 (-0.102, 0.066)	0.350 (0.267, 0.433)	**0.083 (-0.045, 0.211)**	0.005 (0.001, 0.009)

Abbreviations: MR, Mendelian randomization; UV, univariable; MV, multivariable; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TG, triglycerides.
[a] Alleles orientated for all genetic associations with respect to the risk increasing allele for HDL-C, LDL-C or triglycerides.

4.3.3 Summary

In this Section, we applied the methods discussed in Section 4.2 to published summary level data on HDL-C, LDL-C, triglycerides, and CHD risk. When the genetic variants were orientated with respect to the HDL-C risk increasing alleles, the multivariable MR-Egger model containing HDL-C, LDL-C and triglycerides suggested that HDL-C was not causally associated with CHD risk, but there was evidence of residual pleiotropy and/or the InSIDE assumption being violated. The analysis highlighted the sensitivity of the intercept term in the multivariable MR-Egger model to changes in the orientation of the genetic variants. In the next Section, we perform a simulation study based on the applied example considered in this Section to compare the performances of the multivariable IVW, univariable MR-Egger and multivariable MR-Egger methods.

4.4 Simulation study

In order to assess the merits of using multivariable MR-Egger over multivariable IVW and univariable MR-Egger in realistic settings, we perform a simulation study. The code used for the simulation study was written and performed by Jessica Rees in RStudio version 3.5.3 [86].

Univariable and multivariable MR-Egger will be compared with respect to the consistency of the causal estimates and statistical power to detect the causal effect. The setup of the simulation study corresponds to the applied example in Section 4.3 and will be considered under two broad scenarios: (1) $\boldsymbol{\beta_{X_k}}$ are generated independently for all $k = 1, 2, \ldots, K$; and (2) $\boldsymbol{\beta_{X_k}}$ are correlated for all $k = 1, 2, \ldots, K$.

We simulated summarized level data for 185 genetic variants (value taken from the applied example in Section 4.3) indexed by $j = 1, 2, \ldots, J$ for three risk factors (X_1, X_2, X_3) and an outcome Y from the following data-generating model:

$$\begin{aligned}
\begin{pmatrix} \beta_{X_{1j}} \\ \beta_{X_{2j}} \\ \beta_{X_{3j}} \end{pmatrix} &\sim \mathcal{N}_3 \left(\begin{pmatrix} 0.08 \\ 0.03 \\ -0.05 \end{pmatrix}, \begin{pmatrix} {\sigma_1}^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & {\sigma_2}^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & {\sigma_3}^2 \end{pmatrix} \right), \\
\beta_{Yj} &= \alpha'_j + \theta_1|\beta_{X_{1j}}| + \theta_2\beta_{X_{2j}} + \theta_3\beta_{X_{3j}} + \epsilon_j\,, \\
\epsilon_j &\sim \mathcal{N}(0, 1)\,, \\
\alpha'_j &\sim \mathcal{N}(\mu, 0.004)\,. \qquad (4.5)
\end{aligned}$$

The primary objective was to estimate θ_1, with the causal effects set to: $\theta_1 = 0$ (null causal effect) or $\theta_1 = 0.3$ (positive causal effect); $\theta_2 = 0.1$; and $\theta_3 = -0.3$. These causal effects were chosen to ensure the direction and magnitude of effects differed across the three risk factors. The data were simulated to consider the following four scenarios:

1. No pleiotropy ($\alpha'_j = 0$ for all j), InSIDE assumption automatically satisfied;
2. Balanced pleiotropy ($\mu = 0$), InSIDE assumption satisfied;
3. Directional pleiotropy ($\mu = 0.01, 0.05$ or 0.1), InSIDE assumption satisfied;
4. Directional pleiotropy ($\mu = 0.01, 0.05$ or 0.1), InSIDE assumption violated.

The values for μ and the mean values of the genetic associations with the risk factors in Equation 4.5 were chosen to ensure the magnitude of the effects were comparable.

When the InSIDE assumption for multivariable MR-Egger was satisfied, α'_j and $\beta_{X_{1j}}$ were drawn from independent distributions, and when it was violated they were drawn from a multivariate normal distribution with $\text{cor}(\boldsymbol{\alpha'}, \boldsymbol{\beta_{X_1}}) = 0.3$. The above four scenarios were applied to the simulated data when $\boldsymbol{\beta_{X_k}}$ were generated independently for all k, with the parameters in the covariance matrix set to: $\sigma_1^2 = 0.03$; $\sigma_2^2 = 0.02$; $\sigma_3^2 = 0.04$; and $\rho_{12} = \rho_{13} = \rho_{23} = 0$. The four scenarios were repeated when $\boldsymbol{\beta_{X_k}}$ were correlated for all k with $\rho_{12} = 0.2$, $\rho_{13} = -0.3$ and $\rho_{23} = 0.1$. These values were chosen to provide a range in the strength and direction of the correlation structure, and to make sure they were comparable with $\text{cor}(\boldsymbol{\alpha'}, \boldsymbol{\beta_{X_1}})$ when the InSIDE assumption for multivariable MR-Egger was violated. In total, data were simulated for 32 different choices of parameters.

To ensure the direction of association between G_j and X_1 was the same for all j variants, the absolute value of the genetic associations with X_1 ($|\beta_{X_{1j}}|$) were used to generate β_{Y_j} (4.5). It was assumed that $\beta_{X_{kj}}$ (for all k) and β_{Y_j} had the same reference allele and the variants were uncorrelated. The multivariable IVW, univariable MR-Egger and multivariable MR-Egger methods were applied to the simulated datasets. The weights for the multivariable IVW and multivariable MR-Egger are given by equation 4.6, while equation 4.7 contains the weights for univariable MR-Egger:

$$\text{se}(\beta_{Y_j})^{-2} = \left({\epsilon_j}^2 + {\sigma_{\alpha'}}^2\right)^{-1}, \tag{4.6}$$

$$\text{se}(\beta_{Y_j})^{-2} = \left({\epsilon_j}^2 + {\sigma_{\alpha'}}^2 + {\theta_2}^2{\sigma_2}^2 + {\theta_3}^2{\sigma_3}^2\right)^{-1}. \tag{4.7}$$

Since the data–generating model produces summary level data for each risk factor directly, it was not possible to estimate the F-statistic or I^2 statistic without making additional assumptions. For a two-sample Mendelian randomization analysis with summarized data, the F-statistic for each genetic variant j can be approximated by $F_j \approx \hat{\beta}^2_{X_j} / \text{se}(\hat{\beta}_{X_j})^2$, and the formula for the I^2 statistic using summary level data is given in Section 2.6.2. To provide approximate values for both of these summary measures, we require estimates of the standard errors of the genetic associations with the risk factors $\text{se}(\hat{\boldsymbol{\beta}}_{\boldsymbol{X_k}})$. Although the data–generating model made assumptions about the variances of the genetic associations of the risk factors, the standard errors of the genetic associations were not considered. Note that $\text{se}(\hat{\boldsymbol{\beta}}_{\boldsymbol{X_k}})$ are not required to fit the IVW, univariable MR-Egger or multivariable MR-Egger models considered in the simulation study.

We must make additional assumptions to approximate $\text{se}(\hat{\boldsymbol{\beta}}_{\boldsymbol{X_k}})$. If we assume that the genetic associations with the risk factors are measured on the standard deviation

scale, and the associations were estimated from a sample size of 10,000, then this results in a standard error of 0.01. For $\boldsymbol{\beta_{X_1}}$, where the mean value of the genetic associations was 0.08, a standard error of 0.01 would result in an approximate Z-statistic of 8, equivalent to a p-value of 1×10^{-15}. This is relatively close to a Z-statistic of 5.45 that is required for a genome-wide level of significance (p-value $< 5 \times 10^{-8}$).

Assuming that the standard errors of the genetic associations with the three risk factors are 0.01 across the 185 genetic variants, we can estimate the mean F-statistics and I^2 statistics (Table 4.3 and Table 4.4). The I^2 statistics (reported as a %) were close to 100% across the different scenarios, and the mean F-statistics were greater than 200. For multivariable Mendelian randomization analyses, the F-statistic and Sanderson-Windmeijer conditional F-statistic should be provided for each risk factor as a measure of instrument strength [83]. Since the data-generating model did not produce individual level data, we were unable to calculate the conditional F-statistic, although the F-statistics suggested that the genetic variants were strongly associated with the three risk factors.

Table 4.3 Estimates of the mean F-statistic and I^2 statistic (reported as a %) for each risk factor for a null ($\theta_1 = 0$) and positive ($\theta_1 = 0.3$) causal effect where $\boldsymbol{\beta_{X_k}}$ were generated independently for all k.

	$\hat{\beta}_{X_{1j}}$		$\hat{\beta}_{X_{2j}}$		$\hat{\beta}_{X_{3j}}$	
	F-statistic	I^2 statistic	F-statistic	I^2 statistic	F-statistic	I^2 statistic
Null causal effect: $\boldsymbol{\theta_1 = 0}$						
1. No pleiotropy, InSIDE satisfied						
	363.3	99.5	208.8	99.2	425.7	99.6
2. Balanced pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0,0.004)$	364.3	99.5	209.1	99.2	425.6	99.6
3. Directional pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	364.4	99.5	208.9	99.2	425.6	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	363.5	99.5	209.6	99.2	424.9	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	364.0	99.5	209.2	99.2	425.5	99.6
4. Directional pleiotropy, InSIDE violated						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	364.1	99.5	208.8	99.2	425.4	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.4	99.5	208.7	99.2	425.4	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	363.8	99.5	209.1	99.2	425.1	99.6
Positive causal effect: $\boldsymbol{\theta_1 = 0.3}$						
1. No pleiotropy, InSIDE satisfied						
	363.9	99.5	209.2	99.2	424.7	99.6
2. Balanced pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0,0.004)$	363.7	99.5	209.1	99.2	425.0	99.6
3. Directional pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	364.1	99.5	209.0	99.2	425.2	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.3	99.5	208.6	99.2	425.5	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	363.8	99.5	209.1	99.2	424.7	99.6
4. Directional pleiotropy, InSIDE violated						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	363.6	99.5	209.1	99.2	424.8	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.6	99.5	208.9	99.2	424.8	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	364.0	99.5	208.9	99.2	425.9	99.6

Table 4.4 Estimates of the mean F-statistic and I^2 statistic (reported as a %) for each risk factor for a null ($\theta_1 = 0$) and positive ($\theta_1 = 0.3$) causal effect where $\boldsymbol{\beta_{X_k}}$ were correlated for all k.

	$\hat{\beta}_{X_{1j}}$		$\hat{\beta}_{X_{2j}}$		$\hat{\beta}_{X_{3j}}$	
	F-statistic	I^2 statistic	F-statistic	I^2 statistic	F-statistic	I^2 statistic
Null causal effect: $\boldsymbol{\theta_1 = 0}$						
1. No pleiotropy, InSIDE satisfied						
	364.1	99.5	208.5	99.2	424.5	99.6
2. Balanced pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0,0.004)$	363.8	99.5	209.4	99.2	424.4	99.6
3. Directional pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	363.5	99.5	208.8	99.2	424.6	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.3	99.5	209.0	99.2	425.1	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	364.3	99.5	208.9	99.2	424.7	99.6
4. Directional pleiotropy, InSIDE violated						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	364.0	99.5	208.9	99.2	425.0	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.0	99.5	209.4	99.2	425.2	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	364.3	99.5	209.1	99.2	425.2	99.6
Positive causal effect: $\boldsymbol{\theta_1 = 0.3}$						
1. No pleiotropy, InSIDE satisfied						
	364.0	99.5	208.9	99.2	425.2	99.6
2. Balanced pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0,0.004)$	364.0	99.5	208.8	99.2	425.1	99.6
3. Directional pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	363.5	99.5	209.1	99.2	425.5	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	363.9	99.5	209.1	99.2	424.6	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	364.0	99.5	209.1	99.2	425.8	99.6
4. Directional pleiotropy, InSIDE violated						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	364.1	99.5	208.8	99.2	425.3	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.7	99.5	208.8	99.2	425.4	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	363.7	99.5	208.9	99.2	424.5	99.6

4.4.1 Results

The results from the simulation study using 10 000 simulated datasets are presented in Table 4.5 ($\boldsymbol{\beta_{X_k}}$ generated independently) and Table 4.6 ($\boldsymbol{\beta_{X_k}}$ correlated). For each scenario, the mean estimate, the mean standard error, and the statistical power to detect a null or positive causal effect at a nominal 5% significance level are presented in Tables 4.5 and 4.6 for the multivariable IVW, univariable MR-Egger and multivariable MR-Egger methods. For univariable and multivariable MR-Egger, the statistical power of the MR-Egger intercept test is also provided.

$\boldsymbol{\beta_{X_k}}$ generated independently: In scenarios 1 and 2 (no and balanced pleiotropy), estimates from all methods were unbiased, and those from the multivariable IVW method were the most precise. In scenarios 3 and 4 (directional pleiotropy), estimates from the multivariable IVW method were biased, with the magnitude of bias increasing as the average value of $\boldsymbol{\alpha'}$ increased from 0.01 to 0.1. In scenario 3 (InSIDE satisfied), estimates from the univariable and multivariable MR-Egger methods were unbiased, whereas in scenario 4 (InSIDE violated), they were biased. Although the causal estimates for both multivariable IVW and multivariable MR-Egger were biased under scenario 4, the magnitude of bias was less for multivariable MR-Egger, with the exception of when α'_j was generated from $\mathcal{N}(0.01, 0.004)$. Precision and power to detect a causal effect were always better for the multivariable MR-Egger method than univariable MR-Egger, although the univariable MR-Egger method detected directional pleiotropy more often. The average value of $\boldsymbol{\alpha'}$ had no impact on the degree of bias for univariable or multivariable MR-Egger.

$\boldsymbol{\beta_{X_k}}$ correlated: Bias for the multivariable IVW method was present in scenarios 3 and 4 only, as in the independently generated setting. In this setting, the InSIDE assumption for univariable MR-Egger was violated for all four scenarios, resulting in biased point estimates of θ_1. However, the multivariable InSIDE assumption was satisfied for scenarios 1, 2 and 3, and so causal estimates from multivariable MR-Egger were unbiased. When the multivariable InSIDE assumption was violated (scenario 4) the estimates from multivariable MR-Egger were biased, yet the magnitude of bias was less compared with univariable MR-Egger as $|\operatorname{cov}(\boldsymbol{\alpha'}, \boldsymbol{\beta_{X_1}})| < |\operatorname{cov}(\boldsymbol{\alpha}, \boldsymbol{\beta_{X_1}})|$.

Table 4.5 Performance of multivariable IVW, univariable MR-Egger and multivariable MR-Egger with respect to $\hat{\theta}_1$ for a null ($\theta_1 = 0$) and positive ($\theta_1 = 0.3$) causal effect where $\boldsymbol{\beta_{X_k}}$ are generated independently for all k. All tests were performed at the 5% level of significance.

	MV IVW		**UV MR-Egger**			**MV MR-Egger**		
	Mean $\hat{\theta}_1$ (mean SE)	Power, %	Mean $\hat{\theta}_1$ (mean SE)	Power, % Int.	$\hat{\theta}_1$	Mean $\hat{\theta}_1$ (mean SE)	Power, % Int.	$\hat{\theta}_1$
Null causal effect: $\boldsymbol{\theta_1 = 0}$								
1. No pleiotropy, InSIDE satisfied								
	0.000 (0.045)	3.8	-0.002 (0.158)	9.1	4.7	0.000 (0.084)	3.7	4.1
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	-0.001 (0.100)	4.7	-0.001 (0.187)	7.8	4.7	0.000 (0.165)	4.6	4.6
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.041 (0.100)	6.7	-0.003 (0.187)	12.2	4.3	-0.002 (0.165)	5.9	4.5
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.210 (0.100)	55.3	0.002 (0.187)	49.2	4.6	0.002 (0.166)	36.3	4.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.417 (0.102)	97.4	0.000 (0.187)	91.6	4.3	0.001 (0.165)	88.0	4.6
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.074 (0.100)	12.3	0.089 (0.187)	6.7	7.6	0.088 (0.165)	4.3	8.4
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.240 (0.100)	67.2	0.089 (0.187)	34.1	7.8	0.088 (0.165)	21.1	8.8
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.450 (0.101)	98.6	0.088 (0.187)	84.1	7.6	0.088 (0.165)	78.7	8.7
Positive causal effect: $\boldsymbol{\theta_1 = 0.3}$								
1. No pleiotropy, InSIDE satisfied								
	0.300 (0.044)	98.9	0.300 (0.157)	9.3	50.1	0.300 (0.084)	4.3	87.3
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	0.301 (0.100)	84.6	0.303 (0.187)	7.5	38.2	0.302 (0.166)	4.9	46.4
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.343 (0.100)	91.5	0.300 (0.187)	12.8	36.8	0.299 (0.165)	6.0	45.8
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.509 (0.100)	99.7	0.300 (0.188)	50.6	37.3	0.299 (0.166)	37.1	46.1
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.716 (0.102)	100.0	0.300 (0.187)	91.1	37.1	0.299 (0.166)	87.9	46.1
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.374 (0.099)	94.3	0.390 (0.187)	6.6	56.4	0.389 (0.165)	4.6	65.8
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.539 (0.100)	99.8	0.388 (0.187)	34.4	55.6	0.387 (0.165)	21.5	65.5
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.747 (0.101)	100.0	0.383 (0.187)	84.7	55.1	0.384 (0.165)	78.3	65.2

Abbreviations: MR, Mendelian randomization; UV, univariable; MV, multivariable; Int., intercept; SE, standard error; IVW, inverse-variance weighted; InSIDE, Instrument Strength Independent of Direct Effect.

Table 4.6 Performance of multivariable IVW, univariable MR-Egger and multivariable MR-Egger with respect to $\hat{\theta}_1$ for a null ($\theta_1 = 0$) and positive ($\theta_1 = 0.3$) causal effect where $\boldsymbol{\beta_{X_k}}$ were correlated for all k. All tests were performed at the 5% level of significance.

	MV IVW		**UV MR-Egger**			**MV MR-Egger**		
	Mean $\hat{\theta}_1$ (mean SE)	Power, %	Mean $\hat{\theta}_1$ (mean SE)	Power, % Int.	$\hat{\theta}_1$	Mean $\hat{\theta}_1$ (mean SE)	Power, % Int.	$\hat{\theta}_1$
Null causal effect: $\boldsymbol{\theta_1 = 0}$								
1. No pleiotropy, InSIDE satisfied								
	0.000 (0.047)	4.0	0.099 (0.157)	4.3	10.1	0.000 (0.086)	4.4	4.6
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	-0.001 (0.104)	4.7	0.093 (0.187)	4.5	7.4	-0.003 (0.169)	4.6	4.4
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.043 (0.104)	7.0	0.099 (0.187)	5.8	8.0	0.001 (0.169)	5.9	4.8
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.213 (0.105)	52.7	0.095 (0.187)	33.3	7.6	0.000 (0.169)	37.2	4.5
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.426 (0.107)	96.3	0.096 (0.187)	84.5	7.6	-0.001 (0.169)	89.2	4.6
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.062 (0.104)	9.5	0.184 (0.187)	4.6	17.9	0.078 (0.169)	4.7	7.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.235 (0.104)	62.1	0.187 (0.187)	20.5	18.3	0.082 (0.169)	22.3	7.5
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.448 (0.106)	97.9	0.181 (0.187)	73.3	17.8	0.077 (0.169)	80.3	7.2
Positive causal effect: $\boldsymbol{\theta_1 = 0.3}$								
1. No pleiotropy, InSIDE satisfied								
	0.300 (0.047)	98.7	0.395 (0.158)	4.4	70.8	0.299 (0.087)	3.9	86.2
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	0.300 (0.104)	81.5	0.399 (0.187)	4.4	58.0	0.301 (0.169)	4.6	44.4
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.342 (0.104)	89.4	0.395 (0.187)	6.4	57.4	0.301 (0.169)	5.9	44.4
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.513 (0.105)	99.4	0.394 (0.187)	33.0	57.4	0.296 (0.169)	38.0	43.4
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.729 (0.107)	100.0	0.400 (0.187)	83.5	58.2	0.304 (0.169)	88.6	45.5
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.365 (0.104)	92.1	0.489 (0.187)	4.2	74.0	0.382 (0.169)	4.6	63.2
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.535 (0.104)	99.7	0.486 (0.187)	20.3	72.9	0.382 (0.169)	21.1	63.2
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.749 (0.106)	100.0	0.488 (0.187)	72.5	73.4	0.381 (0.169)	79.6	62.8

Abbreviations: MR, Mendelian randomization; UV, univariable; MV, multivariable; Int., intercept; SE, standard error; IVW, inverse-variance weighted; InSIDE, Instrument Strength Independent of Direct Effect.

4.4.2 Summary

In this Section, we have performed a simulation study to investigate the benefits of using multivariable MR-Egger over multivariable IVW and univariable MR-Egger when the risk factors do not have causal effects on each other. The simulation study has highlighted the sensitivity of univariable MR-Egger to the correlation structure of $\boldsymbol{\beta}_{\boldsymbol{X}_k}$, and has demonstrated the merits of using multivariable MR-Egger in terms of the consistency and precision of the causal effect. In the next Section we allow the risk factors to have a causal effect on each other and re-perform the simulation study.

4.5 Causal relationships between risk factors

The simulations performed in Section 4.4 assumed that the effect of each risk factor on the outcome is not mediated through another risk factor. There may be circumstances where causal relationships between risk factors are biologically plausible. Burgess *et al.* [28] illustrated that the multivariable IVW method estimates the direct causal effects (θ_k) of each risk factor on the outcome, irrespective of whether causal relationships between the risk factors exist.

In the applied example (Section 4.3), there may also be deterministic dependencies between the risk factors. LDL-C is rarely measured directly, but is estimated from measurements of total cholesterol, triglycerides and HDL-C via the Friedewald equation as total cholesterol minus HDL-C minus 0.2 times triglycerides (assuming all measurements are in mg/dL) [81]. It has previously been shown that the coefficient for LDL-C is the same as the coefficient for non-HDL-C (calculated as total cholesterol minus HDL-C) in a regression model including HDL-C and triglycerides (see Appendix 2 in the paper by Di Angelantonio *et al.* [119]). However, the coefficient for triglycerides will change, as the non-HDL-C measure contains more triglycerides than the LDL-C measure. Hence, in the case that there are deterministic relationships between the risk factors, effect estimates may change as the choice of risk factors varies due to their interpretation as direct effects conditional on other risk factors in the regression model.

To investigate the behaviour of the multivariable MR-Egger method when associations between risk factors exist, the following simulation study was performed when X_2 was dependent on X_1.

4.5.1 Simulation study

We assume that X_2 is causally dependent on X_1, and the causal effect of X_1 on X_2 is γ (Figure 4.3). The total causal effect of X_1 on Y is $\theta_1 + \gamma\theta_2$; consisting of the direct effect (θ_1) and the indirect effect via X_2 ($\gamma\theta_2$).

The simulations outlined in Section 4.4 were repeated with the second line in the data-generating model (Equation 4.5) replaced with:

$$\beta_{Y_j} = \alpha'_j + \theta_1|\beta_{X_{1j}}| + \theta_2(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|) + \theta_3\beta_{X_{3j}} + \epsilon_j\,.$$

The causal effect of X_1 on X_2 (γ) was set to 0.5. All other parameters were the same as the original simulation study. $|\beta_{X_{1j}}|$, $(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|)$, and $\beta_{X_{3j}}$ were the covariates

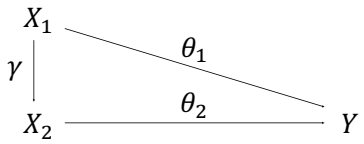


Fig. 4.3 Directed acyclic graph illustrating the causal relationships between the two risk factors X_1 and X_2, and outcome Y. The causal effect of X_1 on X_2 is γ, and the direct causal effect of the risk factor X_k on the outcome Y is θ_k. The total causal effect of X_1 on Y is $\theta_1 + \gamma\theta_2$; consisting of the direct effect (θ_1) and the indirect effect via X_2 ($\gamma\theta_2$).

included in the multivariable IVW and multivariable MR-Egger models. Note that the functional relationship between X_1 and X_2 induces a correlation structure between the covariates $|\beta_{X_{1j}}|$ and $(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|)$ included in the multivariable models, even when $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ are generated independently. To account for the additional uncertainty in β_{Y_j}, the weights for univariable MR-Egger are given by equation 4.8, while the weights for multivariable IVW and multivariable MR-Egger were the same as the original simulation study (equation 4.6).

$$\mathrm{se}(\beta_{Y_j})^{-2} = ({\epsilon_j}^2 + {\sigma_{\alpha'}}^2 + {\theta_2}^2{\sigma_2}^2 + (\theta_2\gamma)^2{\sigma_1}^2 + 2\theta_2\gamma\rho_{12}\sigma_1\sigma_2 + {\theta_3}^2{\sigma_3}^2)^{-1}. \qquad (4.8)$$

Results

The results from the simulations that included a causal relationship between X_1 and X_2, using 10 000 simulated datasets, are presented in Table 4.7 ($\boldsymbol{\beta_{X_k}}$ generated independently, with the functional relationship between X_1 and X_2 inducing a correlation structure between $|\beta_{X_{1j}}|$ and $(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|)$) and Table 4.8 ($\boldsymbol{\beta_{X_k}}$ correlated).

$\boldsymbol{\beta_{X_k}}$ generated independently, with a correlation structure between the covariates $|\beta_{X_{1j}}|$ and $(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|)$: In scenarios where there was no bias in the original set of simulations, the multivariable IVW and multivariable MR-Egger methods consistently estimated the direct effect of X_1 on Y (θ_1), whilst the univariable MR-Egger method consistently estimated the total causal effect of X_1 on Y ($\theta_1 + \gamma\theta_2$). Bias for the multivariable IVW method was present in scenarios 3 and 4 only, as in the original simulation study (Tables 4.5 and 4.6). Compared to the results in Table 4.5, precision and power to detect a causal effect were reduced for the multivariable IVW and multivariable MR-Egger methods. This reduction in power may be due to the correlation structure between $|\beta_{X_{1j}}|$ and $(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|)$, and the multivariable models conditioning on a mediator. Univariable and multivariable MR-Egger methods produced biased estimates of the total and direct causal effects in scenario 4 (InSIDE

violated) only. Unlike the original simulation study, precision and power to detect a causal effect were always better for the univariable MR-Egger method.

$\boldsymbol{\beta_{X_k}}$ correlated: The multivariable IVW and multivariable MR-Egger methods estimated the direct effect of X_1 on Y, as in the independently generated setting. As with the original simulations (Tables 4.5 and 4.6), the InSIDE assumption for univariable MR-Egger was violated for all four scenarios, resulting in biased point estimates. However, as with the original simulation study, the multivariable InSIDE assumption was satisfied for scenarios 1,2 and 3, and so causal estimates from multivariable MR-Egger were unbiased. There was a more noticeable reduction in the precision and power to detect a causal effect for the multivariable IVW and multivariable MR-Egger methods under the correlated setting.

Table 4.7 Performance of multivariable IVW, univariable MR-Egger and multivariable MR-Egger with respect to $\hat{\theta}_1$ for a null ($\theta_1 = 0$) and positive ($\theta_1 = 0.3$) causal effect where $\boldsymbol{\beta_{X_k}}$ are generated independently for all k (with a correlation structure between the covariates $|\beta_{X_{1j}}|$ and $(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|)$), with a causal effect of X_1 on X_2 ($\gamma = 0.5$). All tests were performed at the 5% level of significance.

	MV IVW		**UV MR-Egger**			**MV MR-Egger**		
	Mean $\hat{\theta}_1$ (mean SE)	Power, %	Mean $\hat{\theta}_1$ (mean SE)	Power, % Int.	$\hat{\theta}_1$	Mean $\hat{\theta}_1$ (mean SE)	Power, % Int.	$\hat{\theta}_1$
Null causal effect: $\boldsymbol{\theta_1 = 0}$								
1. No pleiotropy, InSIDE satisfied								
	0.000 (0.057)	3.5	0.051 (0.158)	8.9	5.8	0.001 (0.090)	4.5	4.2
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	0.001 (0.127)	4.4	0.049 (0.187)	7.6	5.6	0.001 (0.178)	4.6	4.2
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.041 (0.127)	6.0	0.049 (0.187)	12.3	5.4	0.000 (0.178)	5.8	4.8
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.195 (0.128)	34.4	0.048 (0.187)	50.1	5.3	-0.001 (0.178)	36.6	4.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.393 (0.130)	82.3	0.052 (0.187)	91.4	5.6	0.002 (0.178)	88.4	4.7
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.076 (0.127)	9.8	0.138 (0.187)	6.4	11.6	0.088 (0.178)	4.3	7.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.231 (0.127)	45.2	0.137 (0.187)	34.4	11.9	0.088 (0.178)	21.7	8.2
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.426 (0.129)	88.3	0.141 (0.187)	83.7	11.9	0.089 (0.178)	78.2	8.1
Positive causal effect: $\boldsymbol{\theta_1 = 0.3}$								
1. No pleiotropy, InSIDE satisfied								
	0.301 (0.057)	96.3	0.353 (0.158)	9.3	62.3	0.301 (0.090)	3.9	84.6
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	0.298 (0.127)	65.4	0.350 (0.187)	7.4	47.8	0.298 (0.178)	4.4	41.2
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.338 (0.127)	75.5	0.352 (0.187)	11.8	48.3	0.300 (0.178)	6.1	41.1
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.494 (0.128)	95.2	0.348 (0.188)	49.2	46.9	0.298 (0.179)	36.8	40.3
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.689 (0.130)	99.6	0.347 (0.188)	91.5	47.1	0.296 (0.178)	88.2	39.6
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.375 (0.127)	82.6	0.440 (0.187)	6.6	65.7	0.390 (0.178)	4.7	60.1
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.530 (0.128)	97.0	0.438 (0.187)	34.7	65.5	0.386 (0.178)	21.7	59.9
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.728 (0.129)	99.7	0.441 (0.187)	83.6	65.8	0.390 (0.178)	78.5	60.1

Abbreviations: MR, Mendelian randomization; UV, univariable; MV, multivariable; Int., intercept; SE, standard error; IVW, inverse-variance weighted; InSIDE, Instrument Strength Independent of Direct Effect.

Table 4.8 Performance of multivariable IVW, univariable MR-Egger and multivariable MR-Egger with respect to $\hat{\theta}_1$ for a null ($\theta_1 = 0$) and positive ($\theta_1 = 0.3$) causal effect where $\boldsymbol{\beta_{X_k}}$ are correlated for all k, with a causal effect of X_1 on X_2 ($\gamma = 0.5$). All tests were performed at the 5% level of significance.

	MV IVW		**UV MR-Egger**			**MV MR-Egger**		
	Mean $\hat{\theta}_1$ (mean SE)	Power, %	Mean $\hat{\theta}_1$ (mean SE)	Power, % Int.	$\hat{\theta}_1$	Mean $\hat{\theta}_1$ (mean SE)	Power, % Int.	$\hat{\theta}_1$
Null causal effect: $\boldsymbol{\theta_1 = 0}$								
1. No pleiotropy, InSIDE satisfied								
	0.000 (0.062)	4.1	0.146 (0.158)	3.9	15.6	0.000 (0.097)	4.0	4.0
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	0.000 (0.137)	4.5	0.146 (0.188)	4.1	11.9	0.000 (0.190)	4.6	4.7
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.041 (0.137)	5.7	0.151 (0.187)	5.4	12.8	0.003 (0.189)	5.7	4.4
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.209 (0.138)	34.2	0.148 (0.187)	32.8	12.6	0.000 (0.190)	36.9	4.7
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.422 (0.140)	82.2	0.151 (0.188)	83.0	12.9	0.004 (0.190)	89.0	4.8
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.053 (0.137)	6.2	0.235 (0.188)	4.3	25.7	0.069 (0.189)	4.9	6.4
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.218 (0.137)	37.2	0.235 (0.188)	20.3	26.4	0.067 (0.189)	21.8	6.7
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.429 (0.139)	84.3	0.238 (0.188)	71.3	26.7	0.071 (0.189)	79.2	6.6
Positive causal effect: $\boldsymbol{\theta_1 = 0.3}$								
1. No pleiotropy, InSIDE satisfied								
	0.299 (0.062)	94.7	0.446 (0.158)	4.1	79.7	0.300 (0.096)	4.0	81.3
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	0.301 (0.137)	60.5	0.445 (0.187)	4.5	66.6	0.300 (0.189)	4.6	37.0
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.339 (0.137)	69.9	0.443 (0.188)	5.7	66.1	0.296 (0.190)	6.0	36.1
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.510 (0.138)	94.2	0.449 (0.188)	32.6	67.7	0.302 (0.190)	37.3	37.2
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.715 (0.140)	99.2	0.445 (0.187)	83.4	66.9	0.298 (0.189)	89.4	36.8
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.353 (0.137)	73.1	0.534 (0.188)	4.4	79.4	0.367 (0.189)	4.6	50.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.519 (0.138)	95.1	0.534 (0.188)	20.3	79.6	0.366 (0.190)	21.7	50.5
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.728 (0.139)	99.5	0.533 (0.188)	72.5	79.6	0.368 (0.189)	80.1	51.0

Abbreviations: MR, Mendelian randomization; UV, univariable; MV, multivariable; Int., intercept; SE, standard error; IVW, inverse-variance weighted; InSIDE, Instrument Strength Independent of Direct Effect.

4.5.2 Summary

In this Section, we have re-performed the simulation study in Section 4.4 to consider the effect of introducing a causal effect between the risk factors in the data-generating model. The multivariable methods estimated the direct effect of the risk factor on the outcome, whilst univariable MR-Egger estimated the total causal effect. In comparison to Section 4.4, the precision and power to detect the causal effect were reduced for the multivariable IVW and multivariable MR-Egger methods.

4.6 Discussion

In this Chapter, we have extended univariable MR-Egger to the multivariable setting and outlined the assumptions required to obtain consistent causal estimates in the presence of directional pleiotropy. Multivariable MR-Egger should be viewed as a sensitivity analysis to provide robustness against both measured and unmeasured pleiotropy, and to strengthen the evidence from the primary Mendelian randomization analysis. If the causal estimate from multivariable MR-Egger is substantially different from the estimate obtained in the primary analysis, then further investigation into the causal finding and the potential for pleiotropy is required.

The simulation studies (Sections 4.4 and 4.5) have highlighted the benefits of using multivariable MR-Egger over its univariable counterpart. This is particularly true when the associations of the genetic variants with the risk factor of interest are associated with genetic associations with at least one of the risk factors (measured pleiotropy). Under this scenario, the InSIDE assumption for univariable MR-Egger is likely to be violated, leading to biased causal estimates. Multivariable MR-Egger will, however, produce consistent causal estimates if the InSIDE assumption for multivariable MR-Egger is satisfied. Although the estimates from univariable and multivariable MR-Egger are asymptotically the same when genetic associations with each risk factor are all independent, multivariable MR-Egger may have greater power to detect a causal effect when the InSIDE assumption is satisfied. Given these advantages, and the sensitivity of the multivariable IVW method to directional pleiotropy, we believe that multivariable MR-Egger should be considered as an important sensitivity analysis for a Mendelian randomization study.

4.6.1 Multivariable by design, or multivariable as a sensitivity analysis?

There are two possible scenarios where multivariable MR-Egger may be used as a sensitivity analysis: either the primary analysis is considered to be multivariable by design, or a multivariable framework is only considered as part of the sensitivity analysis. The first case should be motivated by biological evidence where the set of risk factors are known to be associated with common genetic variants, such as lipid fractions. Under this scenario, multivariable IVW should be used as the primary analysis method with multivariable MR-Egger providing robustness against directional pleiotropy as a sensitivity analysis.

In the second scenario, where there is a lack of biological evidence to suggest a multivariable framework, univariable IVW would generally be considered as the primary analysis method and univariable MR-Egger as the main sensitivity analysis. However, if the genetic variants are associated with other risk factors, multivariable MR-Egger could also be used as a sensitivity analysis as its assumptions are more likely to be satisfied, and it may have greater power to detect a causal effect than univariable MR-Egger. An example of the use of multivariable Mendelian randomization as a sensitivity analysis is a Mendelian randomization study on plasma urate concentrations and CHD risk [120]. To account for measured and unmeasured pleiotropic associations of the genetic variants, the authors performed the multivariable IVW and univariable MR-Egger methods as sensitivity analyses. This investigation may have benefited from performing the multivariable MR-Egger method to simultaneously account for both measured and unmeasured pleiotropic effects.

4.6.2 InSIDE assumption and orientation of genetic variants

The validity of multivariable MR-Egger and its ability to estimate consistent causal effects is dependent upon the InSIDE assumption being satisfied. Whilst it is not possible to determine whether the InSIDE assumption has been violated, we believe it is more likely to hold for multivariable MR-Egger then univariable MR-Egger. When the $\boldsymbol{\beta_{X_1}}$ parameters are correlated with at least one of the sets of $\boldsymbol{\beta_{X_k}}$ parameters for $k = 2, 3, \ldots, K$, the InSIDE assumption for univariable MR-Egger is automatically violated and causal estimates from the method will be inconsistent. Under this setting, the InSIDE assumption for multivariable MR-Egger will be satisfied if all of the parameters $\boldsymbol{\beta_{X_k}}$ $(k = 1, 2, \ldots, K)$ are included in the model. However, it is possible that the remaining direct effect for multivariable MR-Egger contains an unmeasured component that is correlated with one of the $\boldsymbol{\beta_{X_k}}$ $(k = 1, 2, \ldots, K)$ parameters, and the InSIDE assumption for multivariable MR-Egger will be violated. Hence, as highlighted in Section 4.2.5, there is no guarantee that the InSIDE assumption will be more easily satisfied for multivariable MR-Egger than univariable MR-Egger.

The recommendation of orientating the genetic variants in multivariable MR-Egger to the risk factor-increasing or risk factor-decreasing allele for the risk factor of interest may be considered arbitrary. While we accept this limitation, we would argue it brings consistency to the results. This recommendation may result in the analysis being performed up to K times to obtain the causal estimates for all K risk factors. The orientation of the genetic variants will also affect the interpretation of the direct effect, thereby altering the InSIDE assumption. This may result in the MR-Egger

intercept estimate varying between different orientations. This was seen in the applied example (Section 4.3) where the intercept term was non-significant when the alleles were orientated with respect to LDL-C, and significant when orientated with respect to HDL-C and triglycerides.

4.6.3 Linearity and homogeneity assumptions

Throughout this Chapter we have assumed linearity and homogeneity (no effect modification) of the causal effects of the risk factors on the outcome, and of the associations between the genetic variants with the risk factors and with the outcome. Although linearity and homogeneity are strong assumptions, the effect of genetic variants on the risk factor and outcome tend to be limited to a small range, which may make the assumptions of linearity and homogeneity more reasonable in a Mendelian randomization analysis. Note that the assumption of linearity is particularly important in the multivariable setting if there are mediators, as the mediating effects will only cancel out if the associations are linear.

The work presented in this Chapter has focused on effect estimation. Whilst the primary motivation of Mendelian randomization may be the estimation of causal parameters, the estimates themselves can also be used to test the null hypothesis of whether the risk factor is causally associated with the outcome. If the assumptions of linearity and homogeneity are violated then the methods discussed in this Chapter still provide a valid test for the null hypothesis [121], but the estimates will not have a literal interpretation [122].

As highlighted in Sections 2.3.3 and 4.3, the linearity assumption will be violated if the outcome is binary. Under this scenario, the estimates obtained from the Mendelian randomization analysis will be approximations of the causal effects due to the non-collapsibility of the odds ratio. Our justification for extending MR-Egger to the multivariable setting was made under the assumption of linearity, and the simulation study did not consider the impact of a binary outcome as the summary level data was generated directly. This could be viewed as a limitation to the work presented in this Chapter.

The multivariable models have assumed that the risk factors do not have causal effects on each other. The additional simulation study in Section 4.5 has illustrated that the multivariable MR-Egger method estimates the direct causal effects of the risk factors on the outcome, irrespective of whether the risk factors are causally related. There was, however, a reduction in precision and power to detect the causal effect for multivariable MR-Egger when a causal relationship between the risk factors was present.

This reduction in power was anticipated since the multivariable models condition on the mediator along a causal pathway, which is known to decrease power to detect a causal effect [123]. When the risk factors were causally related, univariable MR-Egger will produce consistent causal estimates of the total effect if the InSIDE assumption for univariable MR-Egger is satisfied.

4.6.4 Implication for future research

The paper by Helgadottir *et al.* [114] highlights the importance and need to develop sensitivity analyses for multivariable Mendelian randomization. This is particularly relevant given the recent advances in high-throughput phenotyping which has led to the introduction of '-omics' data such as metabolomics, genomics, and proteomics [124]. Genome-wide analyses of high-dimensional '-omics' data are becoming more popular [125, 126], yet few Mendelian randomization analyses have been performed using these datasets [116]. As summarized data from large consortia become more accessible, the opportunities to use Mendelian randomization on high-dimensional datasets will only increase. Methods such as multivariable MR-Egger will be valuable to investigate the causal effects of multiple related phenotypes with shared genetic predictors.

Bowden *et al.* [49] have shown that uncertainty in the associations between the genetic variants and the risk factor in univariable MR-Egger can lead to attenuation towards the null when a causal effect exists between the risk factor and the outcome in two-sample Mendelian randomization. This attenuation is approximately equal to the I^2 statistic from meta-analysis of the weighted associations with the exposure $\hat{\beta}_{Xj}\,\mathrm{se}(\hat{\beta}_{Yj})^{-1}$, with standard errors $\mathrm{se}(\hat{\beta}_{Xj})\,\mathrm{se}(\hat{\beta}_{Yj})^{-1}$ [49]. Since the estimated mean I^2 statistics for the simulation study in Section 4.4 were close to 100%, there was no substantial bias in the causal estimates due to uncertainty in the genetic associations with the risk factor(s) for either the univariable or multivariable MR-Egger methods. However, it is unclear whether uncertainty in the genetic associations with the risk factors would always lead to the attenuation of the causal estimates for the multivariable MR-Egger method in the two-sample setting.

It was not possible to estimate the Sanderson-Windmeijer conditional F-statistic in the simulation studies as the data-generating model produced summary level data. Although the F-statistics did suggest that the genetic variants were strongly associated with the risk factors, it was unclear whether the variants were jointly associated with the risk factors. As such, we were unable to investigate the effect the conditional F-statistic had on the results from the multivariable methods.

4.6.5 Correlated genetic variants

The methods discussed in this Chapter have assumed that the genetic variants are uncorrelated (not in linkage disequilibrium). There may, however, be cases where using multiple correlated variants from the same gene region will be more efficient than using uncorrelated variants from different gene regions [121]. If the genetic variants are in partial linkage disequilibrium, and each variant explains independent variation in the risk factor, then the inclusion of these variants will increase the power of the Mendelian randomization study. The precision of a Mendelian randomization study will not increase, however, if the variants are perfectly correlated.

If correlated variants are included in an Mendelian randomization study, using summarized level data, the analysis should account for the correlation structure of the variants. If the correlation of the variants is not taken into consideration, the causal estimate will be too precise and this may lead to inappropriate inferences. To account for the correlation between the genetic variants for the univariable and multivariable IVW methods, we can use generalized weighted linear regression of the genetic associations, where the correlations of the variants are included in the weighting matrix, with the intercept set to zero [28, 121].

If $\Omega_{st} = \text{se}(\hat{\beta}_{Y_s})\,\text{se}(\hat{\beta}_{Y_t})\rho_{st}$, where ρ_{st} is the correlation between variants s and t, then the causal estimate from a weighted generalised linear regression for univariable Mendelian randomization is:

$$\hat{\theta}_{UIC} = (\hat{\beta}_{X_j}^T \Omega^{-1} \hat{\beta}_{X_j})^{-1} \hat{\beta}_{X_j}^T \Omega^{-1} \hat{\beta}_{Y_j}\,,$$

with the standard error of the causal estimate:

$$\hat{\theta}_{UIC} = \sqrt{(\hat{\beta}_{X_j}^T \Omega^{-1} \hat{\beta}_{X_j})^{-1}}\,.$$

Whilst the univariable MR-Egger estimates can be obtained by fitting the same generalized weighted linear regression model, but allowing the intercept term to be estimated, the effect of using correlated genetic variants in the univariable MR-Egger method has not been considered in detail. Further investigation into the impact correlated variants may have on the interpretation of the direct effect, and the InSIDE assumption, must be considered at the univariable level first, and then expanded to multivariable MR-Egger.

4.6.6 Key points from chapter

- In this Chapter, we have extended the MR-Egger method to the multivariable setting to correct for both measured and unmeasured pleiotropy.
- Through theoretical arguments and a simulation study, we have shown that the multivariable MR-Egger method has advantages over its univariable counterpart in terms of plausibility of the assumption needed for consistent causal estimation, and power to detect a causal effect when this assumption is satisfied.
- The multivariable MR-Egger method will be useful to analyse high-dimensional data in situations where the risk factors are highly related and it is difficult to find genetic variants specifically associated with the risk factor of interest (multivariable by design), and as a sensitivity analysis when the genetic variants are known to have pleiotropic effects on measured risk factors.

Chapter 5

Extending Mendelian randomization to a factorial framework to detect interaction effects

5.1 Introduction

An interaction between two risk factors occurs when the effect of one risk factor on the outcome is dependent upon the value of the other risk factor. Interactions on disease outcomes can be used to inform public health policies that aim to reduce the burden of disease in the population. The identification of interaction effects may also help to prioritize treatment in the general population and avoid adverse events to treatment. Interaction effects can be considered in randomized clinical trials (RCT) and epidemiological studies using observational data. RCTs can investigate interaction effects to identify sub-populations where the effect of treatment may be more beneficial or even harmful. Interactions in epidemiological studies can highlight whether the effect of a risk factor on a disease outcome can be considered in isolation, or whether its effect should be considered with another risk factor. However, estimates of interaction effects from epidemiological studies and RCTs may not have a causal interpretation due to residual confounding and non-compliance to randomization.

In this Chapter, we explore the possibility of estimating and detecting interaction effects in factorial Mendelian randomization. Unlike Chapters 3 and 4, this Chapter is not centred on developing methodology that accounts for the inclusion of pleiotropic

genetic variants. As highlighted in Section 1.6, the work presented in this Chapter was motivated by the observation that estimating interaction effects between risk factors may be of interest in Mendelian randomization analyses that include two risk factors. Since we demonstrate that summary level data cannot be used in factorial Mendelian randomization (Section 5.4.1), this Chapter focuses on using individual level data for the risk factor, outcome, and genetic variants ('one–sample' Mendelian randomization).

In Section 5.2, we introduce interaction effects within the context of RCTs, and provide an overview of estimating interaction effects. Section 5.3 reviews the literature on factorial Mendelian randomization and justifies the need for further methodological development in this area. In Section 5.4.1, we extend multivariable Mendelian randomization to the factorial setting using individual level data, and demonstrate that standard summary level data cannot be used to estimate the interaction effect of two risk factors in a factorial Mendelian randomization analysis. A formal framework for detecting interactions between drug treatments by using variants as proxies for pharmacological interventions is provided in Section 5.4.2. Finally, we apply the methods considered in this Chapter to data from UK Biobank to estimate the interaction effect of body mass index and alcohol consumption on systolic blood pressure (Section 5.5).

5.2 Interaction effects

This Section defines statistical interaction effects within the context of factorial RCTs, and outlines the types of analyses that may be performed in a factorial RCT. Estimating interaction effects by using linear regression and IVs is then discussed in detail. Finally, we consider the interpretation of statistical interactions as causal effects.

5.2.1 Factorial randomized clinical trials

A factorial RCT allows for the simultaneous assessment of two or more treatments in a single study. In its simplest form, a 2×2 factorial RCT investigates the effect of two binary treatments A and B on a binary outcome Y (Table 5.1). Participants are randomly allocated to one of four groups: to receive treatment A only (n_{10}); to receive treatment B only (n_{01}); to receive both treatments A and B (n_{11}); or to receive neither treatment A nor B (n_{00}). To increase statistical power, participants are normally randomized to create a 'balanced' study where the numbers of participants in each of the four groups are equal [127].

Table 5.1 Illustration of a 2×2 factorial randomized clinical trial where N participants are randomized to treatment A (absence/presence), treatment B (absence/presence), both treatments A and B, or no treatment.

		Treatment A		
		Absence	Presence	
Treatment B	Absence	n_{00}	n_{10}	n_{1+}
	Presence	n_{01}	n_{11}	n_{2+}
		n_{+1}	n_{+2}	N

We let the probability of the outcome Y when treatment $A = a$ and treatment $B = b$ be represented by:

$$p_{ab} = P(Y = 1|A = a, B = b)\,. \tag{5.1}$$

Using Equation 5.1, the statistical interaction of treatment A and treatment B on Y can be measured on the additive scale by:

$$I_+ = p_{11} - p_{10} - p_{01} + p_{00}\,.$$

If $I_+ > 0$, then the effect of both treatments on the outcome is greater than the sum of the individual effects of the treatments, and the interaction is said to be 'positive' or 'super-additive'. If $I_+ < 0$, then the effect of both treatments on the outcome is less than the sum of the individual effects of the treatments, and the interaction is said to be 'negative' or 'sub-additive'.

The effects of the treatments on the outcome can also be considered through the relative risks (RRs) or odds ratios (ORs). Let $RR_{ab} = \frac{p_{ab}}{p_{00}}$ represent the risk of the outcome when $A = a$ and $B = b$ compared to the risk of the outcome in the reference group (no treatment for both A and B), then the statistical interaction of treatment A and treatment B on Y can be measured on the multiplicative scale by:

$$I_\times = \frac{RR_{11}}{RR_{10}RR_{01}} = \frac{p_{11}p_{00}}{p_{10}p_{01}}\,.$$

If $I_\times > 1$, then the effect of both risk factors on the outcome is greater than the product of the individual effects of the risk factors, and if $I_\times < 1$, then the effect of both risk factors on the outcome is less than the product of the individual effects of the risk factors.

Statistical interactions can be assessed by including a product term between the treatments in the regression model. For a linear regression model, this product term

represents an additive statistical interaction, and for a logistic regression model it represents a statistical multiplicative interaction. Statistical interactions may be present on the additive scale but not on the multiplicative scale, and vice-versa [128].

Types of analyses

For a 2×2 factorial RCT, the effects of randomization on the outcome can be assessed by considering the treatments individually, and in combination with each other; allowing for the interaction between the treatments to be investigated. We assume that there is full compliance to randomization, so that the effects of randomization on the outcome from an intention to treat (ITT) analysis are equivalent to the average treatment effects (ATE).

If no interaction effect between the treatments is assumed in a factorial RCT, then the analysis is performed 'at the margins' [129]. Under this approach, the effectiveness of treatment A is considered by comparing the outcome among the participants randomized to treatment A (n_{+2}) with those not randomized to treatment A (n_{+1}), and the effectiveness of treatment B is considered by comparing the outcome in n_{2+} with the outcome in n_{1+}. A multivariable regression model of the outcome with the two treatments and any other necessary covariates, such as baseline characteristics, can be fitted to estimate the treatment effects [127]. For a continuous outcome measure, adjusting for the effect of each treatment will help to improve efficiency, particularly if the study is 'unbalanced' [127]. If the true interaction effect between the treatments is non-zero, then the estimates from this analysis will be biased [129].

When an interaction effect between the treatments is considered, the analysis can be carried out 'inside the table' by treating each group in Table 5.1 as a separate study arm [130, 129]. Under this approach, the effectiveness of treatment A is considered by comparing the outcome in those randomized to receive treatment A only (n_{10}) with those randomized to receive neither treatment A nor B (n_{00}), and the effectiveness of treatment B is considered by comparing the outcome in n_{01} with the outcome in n_{00}. To estimate the statistical interaction effect, the same multivariable regression model described above would be fitted with the addition of a product term between the treatments. Although this type of analysis may be less efficient than assuming there is no interaction effect, the estimates will be unbiased whether the true interaction is zero or not [130].

5.2.2 Interaction effects using observational data

We now consider estimating statistical interaction effects using observational data. First we discuss mean centring continuous variables before generating the product term to improve efficiency. Issues with residual confounding are then highlighted, and estimating the interaction effect using IVs discussed.

Mean centring continuous variables

Suppose we have two modifiable continuous risk factors X_1 and X_2, a continuous outcome Y, and unmeasured confounders U_1 and U_2 of the $X_1 - Y$ and $X_2 - Y$ associations. Consider the following linear regression model for $i = 1, \ldots, N$ observations:

$$Y_i = \theta_0 + \theta_1 X_{1i} + \theta_2 X_{2i} + \theta_{12} X_{12i} + \epsilon_i \,, \tag{5.2}$$

where θ_1 and θ_2 are the main effects of X_1 and X_2, X_{12} is the product of X_1 and X_2, θ_{12} is the statistical interaction effect of X_1 and X_2 on Y, and U_1 and U_2 are contained in the error term ϵ. Since X_{12} is a function of the main effect variables X_1 and X_2, it is likely that X_1 and X_2 will be highly correlated with X_{12}, resulting in reduced statistical power to detect the main effects θ_1 and θ_2. The correlation between these variables can be reduced by mean centring the main effect variables, and then regressing Y against $(X_1 - \bar{X_1})$, $(X_2 - \bar{X_2})$ and $(X_1 - \bar{X_1}) \times (X_2 - \bar{X_2})$, where $\bar{X_1}$ and $\bar{X_2}$ are the mean values of X_1 and X_2 [131]. Mean centring the variables has no impact on the estimate or standard error of the interaction term, but should help to improve the precision of the main effects. By mean centring X_1 and X_2, the main effects will represent the marginal effects θ_{1M} and θ_{2M} of X_1 and X_2 on Y.

Since the linear regression model in Equation 5.2 does not adjust for the unmeasured confounders U_1 and U_2, the estimates obtained from the model will be biased. In the Section below, we consider estimating the interaction effect of X_1 and X_2 on Y using IVs.

Instrumental variables

First consider the following set of equations:

$$\begin{aligned} \mathbb{E}[X_1|Z_1, U_1] &= \beta_{01} + \beta_{11} Z_1 + U_1 \,, \\ \mathbb{E}[X_2|Z_2, U_2] &= \beta_{02} + \beta_{12} Z_2 + U_2 \,, \\ \mathbb{E}[Y|X_1, X_2, U_1, U_2] &= \theta_0 + \theta_1 X_1 + \theta_2 X_2 + U_1 + U_2 \,, \end{aligned} \tag{5.3}$$

where Z_1 and Z_2 are IVs (see Section 1.4 for the IV assumptions) for the risk factors and could be used to estimate θ_1 and θ_2 in a TSLS regression model [38]. For θ_1 and θ_2 to be identifiable in a TSLS model, the number of IVs must be equal to or greater than the number of modifiable risk factors, known as the 'order condition for identification' [132]. Equation 5.3 is 'just identifiable' under the order condition as the number of instruments is equal to the number of risk factors.

Now suppose there is a statistical interaction effect of X_1 and X_2 on Y:

$$\mathbb{E}[Y|X_1, X_2, X_{12}, U_1, U_2] = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_{12} X_{12} + U_1 + U_2 \,. \tag{5.4}$$

Although X_{12} is a function of the original risk factors, X_{12} should be considered as an independent risk factor [133]. We now have three risk factors X_1, X_2 and X_{12}, but only two IVs. In order for Equation 5.4 to be identifiable, X_{12} must have its own instrument [133]. X_{12} can be expressed in terms of Z_1 and Z_2:

$$\begin{aligned}\mathbb{E}[X_{12}|Z_1, Z_2, v_1, v_2] &= (\beta_{11} Z_1 + v_1) \times (\beta_{12} Z_2 + v_2) \\ &= \beta_{11}\beta_{12} Z_1 Z_2 + \beta_{11} Z_1 v_2 + \beta_{12} Z_2 v_1 + v_1 v_2 \,,\end{aligned}$$

where $v_1 = \beta_{01} + U_1$, $v_2 = \beta_{02} + U_2$, and $Z_1 Z_2$ is the product of the IVs Z_1 and Z_2. Since Z_1 and Z_2 are IVs for X_1 and X_2, it follows that $\beta_{11}\beta_{12} \neq 0$, and $Z_1 Z_2$ can act as an IV for X_{12}, making Equation 5.4 just identifiable.

Estimates of θ_1, θ_2 and θ_{12} can be obtained through TSLS regression by regressing the risk factors against the IVs (Z_1, Z_2 and $Z_1 Z_2$) to obtain the predicted values $\hat{X}_1$, $\hat{X}_2$ and $\hat{X}_{12}$, and the outcome Y is then regressed against these predicted values. It would be inappropriate to use $\hat{X}_1$ and $\hat{X}_2$ to obtain the predicted values of X_{12} as $\hat{X}_{12} \neq \hat{X}_1 \times \hat{X}_2$. If this approach was taken, then it would be an example of 'forbidden regression' as the non-linear function of the risk factors X_1 and X_2 is replaced with the same non-linear function of the predicted values from the first stage regression, resulting in inconsistent estimates [133].

Suppose that there are two additional IVs Z_3 and Z_4, such that:

$$\begin{aligned}\mathbb{E}[X_1|Z_1, Z_3, U_1] &= \beta_{01} + \beta_{11} Z_1 + \beta_{21} Z_3 + U_1 \,, \\ \mathbb{E}[X_2|Z_2, Z_4, U_2] &= \beta_{02} + \beta_{12} Z_2 + \beta_{22} Z_4 + U_2 \,,\end{aligned}$$

then $Z_1 Z_2$, $Z_1 Z_4$, $Z_3 Z_2$ and $Z_3 Z_4$ could act as IVs for X_{12} in Equation 5.4. In order for the model to be identifiable, not all of these product terms need to be included

as IVs. In fact, the model would be identifiable with just Z_1, Z_2, Z_3 and Z_4, but the estimates of θ_{12} may be imprecise (considered in Section 5.4.1).

As highlighted at the beginning of Section 5.2.2, including a product term between the main effect variables in a linear regression model can result in reduced statistical power to detect the main effects, and we would expect the TSLS regression model to have a similar issue (considered in Section 5.4.1). X_1 and X_2 should therefore be mean centred before the product term is calculated, and the main effects from the TSLS regression model be interpreted as the marginal effects of X_1 and X_2 on Y.

5.2.3 Causal interpretation

Using the notation in Section 5.2.2, we consider the interpretation of the interaction effect of X_1 and X_2 on the outcome.

It is possible for a statistical interaction effect θ_{12} to have a causal interpretation in an epidemiological study if both sets of confounders U_1 and U_2 have been controlled for [128]. For instance, if the study has taken precise measurements of all the confounders in U_1 and U_2, then the interaction could have a causal interpretation if the regression model adjusted for the effects of U_1 and U_2. However, if the study has not correctly identified and measured the variables in U_1 and U_2, then the interaction will not have a causal interpretation due to residual confounding. Alternatively, the estimate of θ_{12} may have a causal interpretation by controlling for the unmeasured confounders either through the study design, such as a 2×2 factorial RCT, or by using instrumental variable methods, such as TSLS regression.

If there is an interaction between the risk factors, and only U_1 has been controlled for through the study design, then the interaction can only be interpreted as 'effect modification' [134]. For example, let X_1 be a randomized drug in a RCT and X_2 be a baseline characteristic. If the effect of the drug on the outcome differs between the strata of X_2 then we can infer that there is effect modification, but we cannot make any inferences on whether the baseline characteristic is causally relevant in the $X_1 - Y$ association.

If both sets of confounders U_1 and U_2 are controlled for, then the interaction can be interpreted as a causal effect and we can make inferences on how the outcome would be effected if we intervened on both X_1 and X_2 [134]. For example, if X_1 and X_2 were randomized treatments in a 2×2 factorial RCT, then the interaction effect of X_1 and X_2 on the outcome Y would have a causal interpretation. Alternatively, if suitable IVs can be identified then the estimate of the interaction effect from a TSLS regression model would have a causal interpretation.

5.2.4 Summary

In this Section, we have introduced the concept of statistical interaction effects within the context of factorial RCTs, and have discussed the analyses considered in a factorial RCT. We have highlighted the benefits of mean centring continuous variables before generating product terms and demonstrated how causal interactions can be estimated in a TSLS regression model. Finally, we discussed the causal interpretation of interaction effects. In the next Section, we review the literature on factorial Mendelian randomization and IV methodology for estimating interaction effects, providing justification for the work presented in this Chapter.

5.3 Extending Mendelian randomization to a factorial framework

The application of a factorial RCT framework to a Mendelian randomization study has been considered under two broad scenarios: a) to estimate interaction effects between exposures on the risk of disease by using genetic variants as predictors of the risk factors; and b) to identify interactions between drug treatments on the risk of disease by using genetic variants as proxies for pharmacological interventions. As motivation for Section 5.4, the application of these two approaches in the literature, and the need for further methodological developments is discussed below.

5.3.1 Genetic variants used as predictors for the risk factors

The conceptual idea of applying a factorial framework to a Mendelian randomization study was first introduced by Davey Smith and Ebrahim [15]. The term 'factorial Mendelian randomization' is credited by the authors to Sheila Bird. The authors hypothesised that genetic variants could be used as IVs to estimate the simultaneous effect of two or more exposures on the risk of disease, and the interaction effects of the exposures on the outcome. Davey Smith and Hemani [135] emphasised the utility of factorial Mendelian randomization by suggesting that an unconfounded estimate of the interaction of obesity and alcohol consumption on the risk of liver disease could be obtained from this type of study design. However, Davey Smith and Ebrahim's suggestion of using Mendelian randomization to estimate the interaction effect between two risk factors has not been considered further in the literature.

Burgess and Thompson [28] suggested that assessing the causal effects of multiple risk factors in a single Mendelian randomization study, referred to as 'multivariable Mendelian randomization' [28], is analogous to performing a factorial RCT (Figure 5.1). Since multivariable Mendelian randomization assumes that there are no interactions between the risk factors, it would be more accurate to compare this study design to a factorial RCT when the analysis is performed 'at the margins' (Section 5.2.1).

Fig. 5.1 Figure taken from the paper by Burgess and Thompson [28] comparing a factorial randomized clinical trial to a multivariable Mendelian randomization study.

Factorial randomized trial		Randomization of A	
		Control	Treatment A
Randomization of B	Control	Incidence under usual care	Incidence under intervention A
	Treatment B	Incidence under intervention B	Incidence under interventions A and B

Factorial Mendelian randomization		Genetic variant 1	
		Genotype AA	Genotype aa
Genetic variant 2	Genotype BB	AABB (LDL-C, TG usual levels)	aaBB (LDL-C much higher, TG slightly higher)
	Genotype bb	AAbb (LDL-C slightly higher, TG much higher)	aabb (LDL-C much higher, TG much higher)

Web Figure 1: Analogy between a factorial randomized trial with two treatments (A and B) and multivariable Mendelian randomization with two genetic variants (1 and 2). The minor allele of genetic variant 1 (a) has a large effect on LDL-C and a modest effect on triglycerides; the minor allele of genetic variant 2 (b) has a large effect on triglycerides and a modest effect on LDL-C.

Abbreviations: LDL-C = low-density lipoprotein cholesterol, TG = triglycerides.

5.3.2 Genetic variants acting as proxies for pharmacological interventions

The concept of using genetic variants as proxies for pharmacological interventions to identify interaction effects between drug treatments was first introduced by Ference *et al.* [33], and the method has been used in additional studies to assess interactions between drug treatments [34, 35]. Ference *et al.* [33] refer to this type of study design as a '2×2 factorial Mendelian randomization study'. The authors make no reference to the concept of 'factorial Mendelian randomization' as described by Davey Smith and Ebrahim [15].

Ference *et al.* [33] outlined their approach of performing a 2×2 factorial Mendelian randomization study by comparing the effect of lowering low density lipoprotein (LDL-C) levels on the risk of coronary heart disease (CHD) by inhibiting the *NPC1L1* gene with ezetimibe, or inhibiting the *HMGCR* gene with statins, or through a combination

of both. Genetic variants associated with LDL-C levels in either gene region were identified, and two externally weighted gene scores were calculated for each gene region, where the reference alleles were the LDL-C lowering allele for each variant. To mimic a 2×2 factorial RCT, the two gene scores were dichotomized to create a 2×2 contingency table (Table 5.2). The gene scores were dichotomized at their median values to ensure the numbers of participants were balanced across the four groups in Table 5.2, where:

- n_{00} represents the reference group, which was considered to be equivalent to receiving no treatment,
- n_{10} are the group of participants with lower LDL-C mediated by *NPC1L1*, which was considered to be equivalent to receiving the ezetimibe treatment only,
- n_{01} are the group of participants with lower LDL-C mediated by *HMGCR*, which was considered to be equivalent to receiving the statin treatment only, and
- n_{11} are the group of participants with lower LDL-C mediated by *NPC1L1* and *HMGCR*, which was considered to be equivalent to receiving both ezetimibe and statin treatments.

The authors performed an 'inside the table' analysis (Section 5.2.1) by fitting separate logistic regression models to each subgroup, with the n_{00} participants used as the reference group for each model. By comparing the three OR estimates from the separate logistic regression models, the authors concluded that there was no evidence of an interaction effect between ezetimibe and statins, arguing that the effect of lowering LDL-C on the risk of CHD mediated by variants in *NPC1L1*, *HMGCR* or both, was approximately the same. No formal statistical testing was applied to the three OR estimates, and the authors did not attempt to estimate the interaction effect of lowering LDL-C levels by inhibiting the *NPC1L1* gene and inhibiting the *HMGCR* gene on the risk of CHD.

Table 5.2 Contingency table created by Ference *et al.* [33] to compare the effect of lowering low densty lipoprotein levels on the risk of coronary heart disease by inhibiting the *NPC1L1* gene with ezetimibe, either alone, or in combination with a statin that inhibits the *HMGCR* gene.

		Gene score for *NPC1L1*, GS_A	
		$\leq med(GS_A)$	$> med(GS_A)$
Gene score for *HMGCR*, GS_B	$\leq med(GS_B)$	n_{00}	n_{10}
	$> med(GS_B)$	n_{01}	n_{11}

5.3.3 Effect of obesity and alcohol consumption on the risk of liver disease

Although the method proposed by Ference *et al.* [33] (Section 5.3.2) is specifically designed for considering drug interactions, Carter *et al.* [136] have applied the method to investigate the interaction effect of obesity and alcohol consumption on liver disease using data from the Copenhagen General Population Study. As highlighted in Section 5.3.1, the idea of using Mendelian randomization to consider this research question was initially proposed by Davey Smith and Ebrahim [15]. Since Carter *et al.* [136] used the method proposed by Ference *et al.* [33], they were unable to estimate the interaction effect of obesity and alcohol consumption on liver disease. Carter *et al.* [136] may have followed the approach outlined by Ference *et al.* [33] as the current Mendelian randomization literature does not cover the estimation of interaction effects between risk factors.

In their study, Carter *et al.* [136] created a weighted gene score for BMI using five genetic variants, and classified participants as having a 'low BMI' or 'high BMI' if their gene score was $\leq$ or $>$ than the median value of the weighted gene score. The rs1229984 variant in the *ADH1B* gene region was used as a proxy for alcohol consumption. Participants were classified as having a 'low alcohol consumption' if they were homozygous or heterozygous for the alcohol decreasing allele, or a 'high alcohol consumption' if they were homozygous for the alcohol increasing allele. Following the method proposed by Ference *et al.* [33], these classifications were used to create four subgroups of participants: 1) low BMI, low alcohol consumption; 2) low BMI, high alcohol consumption; 3) high BMI, low alcohol consumption; and 4) high BMI, high alcohol consumption.

Carter *et al.* [136] used two plasma biomarkers of liver injury and incident cases of liver disease from hospital records as the outcome measurements. Using the participants with a high BMI and high alcohol consumption as the reference group, three separate regression models were fitted to each outcome measurement to estimate the mean differences (two plasma measurements) or ORs (incident liver disease) for the three remaining groups of participants. The estimates and 95% confidence intervals from these three models were compared to consider the direction and overall patterns of association for each outcome measurement. The authors did not estimate the interaction effect between BMI and alcohol consumption on liver disease.

5.3.4 Requirement for further methodological research

We now review the literature on estimating interaction effects in Mendelian randomization, and outline the new material considered in this Chapter. Since methodological developments in Mendelian randomization is heavily linked to the IV literature, a review of relevant IV methods that estimate interaction effects will be considered first.

Instrumental variable analyses

There has been little method development in the IV literature on estimating interaction effects using observational data. However, interactions can be estimated from observational data using TSLS regression as described in Section 5.2.2.

Whilst there has been a substantial amount of method development in the IV literature on non-compliance in RCTs with a single treatment, there is little guidance on how IVs can be used when there is non-compliance in studies where the interaction between treatments is of primary concern [137]. Since non-compliance to randomization in a 2×2 factorial RCT can lead to artificial interaction effects, reduced statistical power, and biased estimates under an ITT analysis [138], developing methodology that can be used in a 2×2 factorial RCT would be beneficial. Blackwell [137], recognising this gap in the literature, has used a compliance framework to provide non-parametric estimators of the local average interaction effect (LAIE) and the local average conditional effects (LACEs) for a 2×2 factorial study design. The estimators of the LAIE and LACEs (local average treatment effect for one treatment with the other treatment fixed) require estimates of the compliance probabilities for the different treatment groups.

The application of Blackwell's method to Mendelian randomization is limited. To estimate the LAIE under Blackwell's [137] method, estimates of the compliance probabilities are required. Whilst it would be possible for these probabilities to be estimated in a 2×2 factorial RCT, it is difficult to conceive how these probabilities could be obtained in a 2×2 factorial Mendelian randomization study. The applicability of Blackwell's method to a Mendelian randomization study is further restricted by the method only considering binary treatments, with each treatment having its own binary IV. Most Mendelian randomization studies use multiple variants as IVs, and the risk factors are typically continuous [139].

Mendelian randomization

As highlighted in Section 5.3.1, the first paper to introduce the idea of estimating interaction effects between risk factors in Mendelian randomization was written by Davey Smith and Ebrahim in 2003. There has been no methodological developments in the Mendelian randomization literature on estimating interaction effects since this paper was published. We will address this gap in the literature by extending the multivariable Mendelian randomization method to the factorial setting by estimating the interaction term between two risk factors. This expansion will be considered for both individual and summary level data. We will also investigate whether the definition of a 'strong instrument' for multivariable Mendelian randomization (Section 2.6.3) is applicable to the factorial setting. Finally, the suitability of using the method proposed by Ference *et al.* [33] to investigate the interaction effect between two risk factors, as done by Carter *et al.* [136], will be investigated.

The application of a factorial framework to a Mendelian randomization study when the genetic variants are used as proxies for drug treatments has only been considered in the context of applied projects [33–35]. There are several methodological issues relating to the implementation of this method that will be addressed in this Chapter. Rather than performing an 'inside the table' analysis and comparing the estimates from the separate regression models with no statistical testing for an interaction effect, we will fit a single multivariable regression model with an interaction term between the gene scores. Although the interpretability of the interaction effect from this model would be limited, as it represents the effect of the genetic variants on the outcome [33], rather than the effect of the treatments on the outcome, the estimate could be used to assess whether there is evidence of an interaction effect. Instead of dichotomizing the gene scores at their median values, the impact of treating the gene scores as continuous variables to increase statistical power will also be investigated. Whether the gene scores are dichotomized or not, the distribution of the gene scores will effect the power of the study to detect the interaction term, and we will consider this in detail. For example, if the genetic variants are rare, we would expect the distribution of the gene scores to be skewed, and the numbers of participants in the 2×2 contingency table to be unbalanced, resulting in reduced statistical power.

Throughout this Chapter, we will continue to make a distinction between a factorial Mendelian randomization study that uses genetic variants as either: a) predictors of the risk factors to estimate the interaction effect between two risk factors on an outcome; or b) proxies for pharmacological interventions to detect interaction effects between drug treatments. For scenario a), we will use the genetic variants as IVs to

estimate the interaction effect between two risk factors, and for scenario b), we will use the genetic variants as proxies for drug treatments to test for an interaction effect between the variants and the outcome. We class these two study types as 'factorial Mendelian randomization' as they both consider the detection and/or estimation of an interaction effect. Since the primary aim of the analysis and role of the genetic variants differ between the two scenarios, the methodological developments for these study types will be considered separately.

5.3.5 Summary

In this Section, we have reviewed the literature on factorial Mendelian randomization, and have found that the study design has been considered under two broad scenarios: a) the genetic variants are used as predictors of the risk factors; and b) the genetic variants are used as proxies for pharmacological interventions. However, there remain significant gaps in the literature on this topic. The work presented in this Chapter will contribute to the factorial Mendelian randomization literature by investigating how interaction effects can be estimated under scenario a), and providing a more formal framework for detecting interaction effects under scenario b).

5.4 Performing factorial Mendelian randomization

This Section is divided into two main parts: 1) investigating methodological challenges of estimating statistical interactions when genetic variants are used as predictors of the risk factors (Section 5.4.1); and 2) detecting interaction effects between pharmacological interventions when genetic variants are used as proxies for drug treatments (Section 5.4.2).

In Section 5.4.1, we expand multivariable Mendelian randomization with two risk factors to the factorial setting by outlining an approach to estimating the interaction effect through TSLS regression using individual level data. The performance of the TSLS regression model, and the suitability of applying Ference *et al's.* [33] method when the variants are used as predictors of the risk factors is considered in a simulation study. Since summary level data is commonly used in Mendelian randomization, we also consider applying a factorial framework to this type of data. Through theoretical arguments and simulations, we show that standard summary level data cannot be used in factorial Mendelian randomization to estimate the interaction effect of two risk factors.

In Section 5.4.2, we provide a formal framework for using genetic variants as proxies for pharmacological interventions to detect statistical interactions between drug treatments, and through simulations, we address the methodological issues outlined in Section 5.3.4 in relation to the method proposed by Ference *et al.* [33].

We found that some of the distributions of the parameter estimates and standard errors in the simulation studies considered in this Section were slightly positively skewed. As such, we present the median values of the parameter estimates and standard errors for all of the simulations performed in this Chapter. All of the code for the simulations were written and performed by Jessica Rees in RStudio version 3.5.3 [86] using the packages *ivpack* [140] and *ggplot2* [141].

5.4.1 Genetic variants used as predictors of risk factors

Suppose we have two risk factors X_1 and X_2, an outcome Y, and unmeasured confounders U_1 and U_2 of the $X_1 - Y$ and $X_2 - Y$ associations. If there is a set of J genetic variants G_j $(j = 1, \ldots, J)$, where each variant satisfies the following IV assumptions for multivariable Mendelian randomization:

- IV1(M): the variant is associated with at least one of the risk factors X_1 or X_2,
- IV2(M): X_1 and X_2 are associated with at least one of the J genetic variants G_j $(j = 1, \ldots, J)$,
- IV3(M): the variant is independent of all unmeasured confounders U_1 and U_2 of the $X_1 - Y$ and $X_2 - Y$ associations, and
- IV4(M): the variant is independent of the outcome Y conditional on the risk factors X_1 and X_2, and confounders U_1 and U_2 [28].

then the causal relationships between the risk factors and outcome can be assessed through a multivariable Mendelian randomization analysis (Figure 5.2). From Figure 5.2, assuming that all of the relationships are linear on the additive scale, and the genetic variants are uncorrelated (not in linkage disequilibrium), the outcome can be modelled as:

$$\mathbb{E}[Y|X_1, X_2, U_1, U_2] = \theta_{0M} + \theta_{1MV}X_1 + \theta_{2MV}X_2 + \zeta_{1MV}U_1 + \zeta_{2MV}U_2\,, \tag{5.5}$$

where θ_{1MV} and θ_{2MV} are the causal effects of the risk factors X_1 and X_2 on the outcome Y, when conditioned on each other and U_1 and U_2, and ζ_{1MV} and ζ_{2MV} are the effects of the unmeasured confounders U_1 and U_2 on the outcome. In Figure 5.2,

we assume that X_1 and X_2 do not have causal effects on each other.

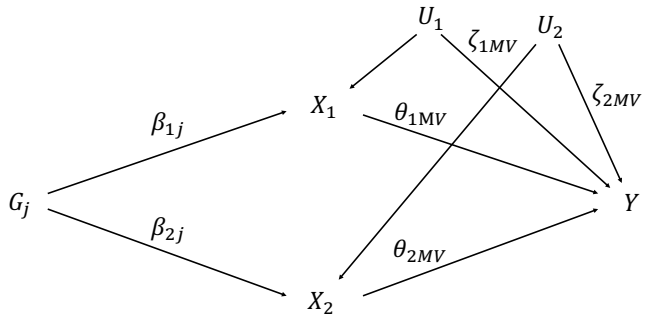


Fig. 5.2 Directed acyclic graph illustrating a multivariable Mendelian randomization framework for a set of genetic variants $\boldsymbol{G}$, two risk factors X_1 and X_2, and outcome Y. The genetic effects of G_j on X_1 and X_2 are β_{1j} and β_{2j} respectively, and the causal effects of the risk factors X_1 and X_2 on the outcome Y is θ_{1MV} and θ_{2MV}. U_1 and U_2 represent the set of unmeasured variables that confound the associations between $X_1 - Y$ and $X_2 - Y$, and ζ_{1MV} and ζ_{2MV} are the effects of the unmeasured confounders U_1 and U_2 on the outcome.

Estimates of the causal effects in Equation 5.5 can be obtained from individual-level data using TSLS regression, with each G_j acting as an IV [28]. The same estimates can also be obtained through the multivariable IVW method [80], where the J genetic associations with Y are regressed against the genetic associations with X_1 and X_2 in a multivariable weighted linear regression model (Section 2.6.3), with the intercept is set to zero:

$$\hat{\beta}_{Y_j} = \theta_{1MV}\hat{\beta}_{X_1 j} + \theta_{2MV}\hat{\beta}_{X_2 j} + \epsilon_{MV_j}, \quad \text{weights} = \text{se}(\hat{\beta}_{Y_j})^{-2}, \tag{5.6}$$

where $\hat{\beta}_{X_1 j}$, $\hat{\beta}_{X_2 j}$ and $\hat{\beta}_{Y_j}$ are estimates of the j genetic associations with X_1, X_2 and Y, and $\text{se}(\hat{\beta}_{Y_j})$ are the standard errors of the genetic associations with Y.

A genetic variant is considered to be a strong instrument in a multivariable Mendelian randomization analysis if: a) the variant is associated with both X_1 and X_2; and b) the variant is jointly associated with X_1 and X_2 [82]. The first condition can be assessed through the F-statistic of the J genetic variants G_j for X_1 and X_2. For the second condition to hold, the variant must be able to predict the values of X_1 after predicting X_2, and this can be assessed through the Sanderson-Windmeijer conditional F-statistic of the J genetic variants G_j for X_1 and X_2 [83] (Section 2.6.3).

Equation 5.5 implies that there is no interaction of X_1 and X_2 on Y. However, we may suspect that a statistical interaction does exist, and the outcome should be

modelled as:

$$\mathbb{E}[Y|X_1, X_2, X_{12}, U_1, U_2] = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_{12} X_{12} + \zeta_1 U_1 + \zeta_2 U_2 \,, \tag{5.7}$$

where X_{12} is the product of the two risk factors $X_1 \times X_2$, and θ_1 and θ_2 are the main causal effects of the risk factors X_1 and X_2 on the outcome Y, when conditioned on each other and U_1 and U_2, and θ_{12} represents the additive statistical interaction effect of X_1 and X_2 on Y. A factorial Mendelian randomization analysis is primarily interested in estimating the interaction effect θ_{12}. Including X_{12} in Equation 5.7 should not invalidate the IV assumptions for multivariable Mendelian randomization if the interaction is accounted for in the analysis using individual level data. The implication of there being an interaction effect when using the summary level data will be discussed towards the end of this Section.

The J genetic variants G_j $(j = 1, \ldots, J)$ consist of three separate sets of genetic variants G_{1j} $(j = 1, \ldots, J_1)$, G_{2j} $(j = 1, \ldots, J_2)$ and G_{cj} $(j = 1, \ldots, J_c)$, where:

- $\boldsymbol{G_1}$ (bold variables represent vectors) are the variants associated with X_1 only, with genetic associations $\boldsymbol{\beta_1}$,
- $\boldsymbol{G_2}$ are the variants associated with X_2 only, with genetic associations $\boldsymbol{\beta_2}$, and
- $\boldsymbol{G_c}$ are the set of cross-over variants that are associated with both X_1 and X_2, with genetic associations $\boldsymbol{\beta_{1_c}}$ and $\boldsymbol{\beta_{2_c}}$ respectively.

The risk factors X_1 and X_2 can be written in terms of $\boldsymbol{G_1}$, $\boldsymbol{G_2}$, and $\boldsymbol{G_c}$ as:

$$\mathbb{E}[X_1|G_1, G_c, U_1] = \beta_{01} + \sum_{j=1}^{J_1} \beta_{1j} G_{1j} + \sum_{j=1}^{J_c} \beta_{1_c j} G_{cj} + U_1 \quad \text{and}$$
$$\mathbb{E}[X_2|G_2, G_c, U_2] = \beta_{02} + \sum_{j=1}^{J_2} \beta_{2j} G_{2j} + \sum_{j=1}^{J_c} \beta_{2_c j} G_{cj} + U_2 \,,$$

where $\boldsymbol{G_1}$ and $\boldsymbol{G_c}$ act as IVs for X_1, and $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$ act as IVs for X_2. Although X_{12} is a function of the risk factors X_1 and X_2, X_{12} should be treated as a separate risk factor, where the products of the IVs for X_1 $(\boldsymbol{G_1} + \boldsymbol{G_c})$ and X_2 $(\boldsymbol{G_2} + \boldsymbol{G_c})$ act as IVs (Section 5.2.2).

Individual level data

Suppose that we have individual level data on X_1, X_2, Y, and J genetic variants G_j ($j = 1, \ldots, J$), and we want to estimate θ_1, θ_2 and θ_{12} in Equation 5.7. To estimate these effects, a TSLS regression model should be applied to the risk factors X_1, X_2 and X_{12}, with $\boldsymbol{G_1}$, $\boldsymbol{G_2}$, $\boldsymbol{G_c}$ and $(\boldsymbol{G_1} + \boldsymbol{G_c}) \times (\boldsymbol{G_2} + \boldsymbol{G_c})$ acting as IVs. To increase statistical power to detect the main effects, the risk factors should be replaced with their mean centred values $(X_1 - \bar{X_1})$, $(X_2 - \bar{X_2})$, and $(X_1 - \bar{X_1}) \times (X_2 - \bar{X_2})$. If the mean centred variables are used in the analysis, then the main effects θ_1 and θ_2 should be interpreted as the marginal effects θ_{1M} and θ_{2M} of X_1 and X_2 on Y.

The number of product terms used as IVs for X_{12} is dependent on the number of cross-over variants J_c. If there are no cross-over variants ($J_c = 0$), then the $\boldsymbol{G_1} \times \boldsymbol{G_2}$ interaction terms could be used as IVs for X_{12}. If $J_c \geq 1$, then $(\boldsymbol{G_1} + \boldsymbol{G_c}) \times (\boldsymbol{G_2} + \boldsymbol{G_c})$ will consist of interaction and quadratic terms. Since we only need three IVs for Equation 5.7 to be identifiable under the order condition (Section 5.2.2), a subset of the product terms in $(\boldsymbol{G_1} + \boldsymbol{G_c}) \times (\boldsymbol{G_2} + \boldsymbol{G_c})$, when $J_c \geq 0$, could act as IVs for X_{12}.

Rather than treating each of the genetic variants as individual IVs, the genetic variants could be combined into two externally weighted gene scores GS_{X_1} and GS_{X_2} for X_1 and X_2:

$$GS_{X_1} = \sum_{j=1}^{J_1} \beta'_{1j} G_{1j} + \sum_{j=1}^{J_c} \beta'_{1_c j} G_{cj} \quad \text{and}$$
$$GS_{X_2} = \sum_{j=1}^{J_2} \beta'_{2j} G_{2j} + \sum_{j=1}^{J_c} \beta'_{2_c j} G_{cj} \,,$$

where β'_1, β'_{1_c}, β'_2, and β'_{2_c} are the genetic associations estimated from an external dataset to act as weights in the gene scores. Note that GS_{X_1} contains information on $\boldsymbol{G_1}$ and $\boldsymbol{G_c}$, whilst GS_{X_2} contains information on $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$. The two gene scores GS_{X_1} and GS_{X_2} and their product $GS_{X_1} \times GS_{X_2}$ could then be used as IVs in the TSLS regression model. Alternatively, the idea proposed by Ference *et al.* [33] to dichotomize the gene scores at their median value could be applied, and the two binary variables and their product be used as IVs in the TSLS regression model. By dichotomizing the gene scores, a contingency table similar to Table 5.2 could be created, where:

- n_{00} are the individuals with $GS_{X_{1i}} \leq med(GS_{X_1})$ and $GS_{X_{2i}} \leq med(GS_{X_2})$, and represents the group of individuals with genetically low measurements for both X_1 and X_2,

- n_{10} are the individuals with $GS_{X_{1i}} > med(GS_{X_1})$ and $GS_{X_{2i}} \leq med(GS_{X_2})$, and represents the group of individuals with genetically high measurements for X_1 and genetically low measurements for X_2,
- n_{01} are the individuals with $GS_{X_{1i}} \leq med(GS_{X_1})$ and $GS_{X_{2i}} > med(GS_{X_2})$, and represents the group of individuals with genetically low measurements for X_1 and genetically high measurements for X_2, and
- n_{11} are the individuals with $GS_{X_{1i}} > med(GS_{X_1})$ and $GS_{X_{2i}} > med(GS_{X_2})$, and represents the group of individuals with genetically high measurements for both X_1 and X_2.

Simulation study

The two risk factors X_1 and X_2 were generated for $i = 1, 2, \ldots, 10\,000$ participants from the following data generating model:

$$\begin{aligned} X_{1i} &= 0.25 + \sum_{j=1}^{J_1} \beta_{1j} G_{1ji} + \sum_{j=1}^{J_c} \beta_{1_c j} G_{cji} + U_{1i} + \epsilon_{1i} \quad \text{and} \\ X_{2i} &= 0.3 + \sum_{j=1}^{J_2} \beta_{2j} G_{2ji} + \sum_{j=1}^{J_c} \beta_{2_c j} G_{cji} + U_{2i} + \epsilon_{2i}\,, \end{aligned} \tag{5.8}$$

where $\boldsymbol{G_1}$ and $\boldsymbol{G_2}$ are the genetic variants associated with X_1 and X_2 respectively, and $\boldsymbol{G_c}$ are the set of cross-over variants that are associated with both X_1 and X_2. The genotypes (0, 1 or 2) were generated independently from binomial distributions $\mathcal{B}(2, MAF_j)$, where MAF_j represents the minor allele frequency (MAF) of the j^{th} genetic variant, and was drawn from a uniform distribution of common genetic variation $\mathcal{U}(0.1, 0.5)$. These bounds for the uniform distribution were chosen to ensure there was a good range in the MAF as these were used to calculate the genetic associations (see below). $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_{1_c}}$ represent the effects of the genetic variants $\boldsymbol{G_1}$ and $\boldsymbol{G_c}$ on X_1, and $\boldsymbol{\beta_2}$ and $\boldsymbol{\beta_{2_c}}$ represent the effects of the genetic variants $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$ on X_2. The genetic associations were initially calculated so that $\boldsymbol{G_1}$ and $\boldsymbol{G_c}$, and $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$, explained $\sigma_1^2 = \sigma_2^2 = 10\%$ of the variance in X_1 and X_2 respectively. Given that genetic variants tend to explain a small proportion of variation in risk factors, we acknowledge that σ_1^2 and σ_2^2 may be considered too large. To ensure that each genetic variant explained the

same amount of variation in the risk factor, we rearranged:

$$\begin{aligned}\text{var}(G_{1j}) &= \sigma_1^2 = 2\beta_{1j}^2 MAF_{1j}(1 - MAF_{1j}) \quad \text{and} \\ \text{var}(G_{2j}) &= \sigma_2^2 = 2\beta_{2j}^2 MAF_{2j}(1 - MAF_{2j})\,,\end{aligned}$$

to calculate the genetic associations:

$$\begin{aligned}\beta_{1j} &= \sqrt{\frac{\sigma_1^2/(J_1 + J_c)}{2 \times MAF_{1j}(1 - MAF_{1j})}}\,, \\ \beta_{c_1 j} &= \sqrt{\frac{\sigma_1^2/(J_1 + J_c)}{2 \times MAF_{cj}(1 - MAF_{cj})}}\,, \\ \beta_{2j} &= \sqrt{\frac{\sigma_2^2/(J_2 + J_c)}{2 \times MAF_{2j}(1 - MAF_{2j})}}\,, \\ \beta_{c_2 j} &= \sqrt{\frac{\sigma_2^2/(J_2 + J_c)}{2 \times MAF_{cj}(1 - MAF_{cj})}}\,.\end{aligned}$$

U_1 and U_2 represent the set of unmeasured confounding variables of the $X_1 - Y$ and $X_2 - Y$ associations. To ensure the unmeasured confounders explained 25% of the variation in the risk factors, U_1 and U_2 were drawn independently from $\mathcal{N}(0, 0.25)$. To fix the variances of X_1 and X_2 to one, the error terms ϵ_1 and ϵ_2 were generated independently from a normal distribution with mean zero, and variances:

$$\sigma_{\epsilon_1}^2 = 1 - \sigma_1^2 - 0.25 \quad \text{and} \quad \sigma_{\epsilon_2}^2 = 1 - \sigma_2^2 - 0.25\,.$$

The outcome Y was generated from:

$$Y_i = \theta_0 + \theta_1 X_{1i} + \theta_2 X_{2i} + \theta_{12} X_{12i} + 0.5U_{1i} + 0.5U_{2i} + \epsilon_{Yi}\,, \tag{5.9}$$

where θ_1 and θ_2 represent the main effects of X_1 and X_2 on Y, and θ_{12} represents the interaction effect of X_1 and X_2 on Y. X_{12} was generated by either: a) multiplying X_1 and X_2; or b) multiplying the mean centred values of the risk factors $(X_1 - \bar{X_1})$ and $(X_2 - \bar{X_2})$, where $\bar{X_1}$ and $\bar{X_2}$ are the mean values of X_1 and X_2. To ensure the risk factors and unmeasured confounders explained less than a third of the variance in the outcome, the error term ϵ_Y was generated from a standard normal distribution $\mathcal{N}(0, 1)$.

TSLS regression models were fitted to either: a) the directly generated values of the risk factors (X_1, X_2, $X_{12} = X_1 \times X_2$); or b) the mean centred values of the risk

factors ($X_1 - \bar{X}_1$, $X_2 - \bar{X}_2$, $X_{12} = (X_1 - \bar{X}_1) \times (X_2 - \bar{X}_2)$). When the risk factors were mean centred, the model estimated the marginal effects θ_{1M} and θ_{2M} of X_1 and X_2 on Y, otherwise θ_1 and θ_2 were estimated. For example, when there were no cross-over variants $J_c = 0$ the marginal effect θ_{M1} is equal to the partial derivative of Equation 5.9 with respect to X_1:

$$\theta_{M1} = \theta_1 + \theta_{12}\mathbb{E}[X_2]\,,$$

and from Equation 5.8:

$$\theta_{M1} = \theta_1 + \theta_{12}\mathbb{E}[0.3 + \sum_{j=1}^{J_2} \beta_{2j} G_{2j} + U_2 + \epsilon_2]\,.$$

Since U_2 and ϵ_2 were generated from normal distributions with means zero, and the mean MAF for $\boldsymbol{G_2}$ will be 0.3 as it was generated from $\mathcal{U}(0.1, 0.5)$, it follows:

$$\theta_{M1} = \theta_1 + 0.3\theta_{12} + J_2\theta_{12}\left(\sqrt{\frac{0.1/J_2}{2 \times 0.3 \times 0.7}} \times 0.3 \times 2\right). \tag{5.10}$$

Under the same argument, the marginal effect for θ_{M2} is:

$$\theta_{M2} = \theta_2 + 0.25\theta_{12} + J_1\theta_{12}\left(\sqrt{\frac{0.1/J_1}{2 \times 0.3 \times 0.7}} \times 0.3 \times 2\right). \tag{5.11}$$

The genetic variants were either treated as individual IVs or as a single instrument in externally weighted gene scores GS_{X_1} and GS_{X_2} for X_1 and X_2. The external weights for the gene scores were based on an independent set of 10 000 individuals, and were produced from the same data generating model used for the main set of participants. Since the same data generating model was used to create the individual level data and the weights for the gene score, GS_{X_1} and GS_{X_2} represent the optimal gene scores. The two gene scores GS_{X_1} and GS_{X_2} were dichotomized at their median values to create two binary variables, and the numbers of participants in the groups n_{00}, n_{10}, n_{01}, and n_{11} were recorded. The following four sets of genetic variants were used as IVs in separate TSLS regression models:

- Model 1 - full set of interactions: the J_1, J_2 and J_c genetic variants used to generate X_1 and X_2, plus the unique interactions and quadratic terms of $(\boldsymbol{G_1} + \boldsymbol{G_c}) \times (\boldsymbol{G_2} + \boldsymbol{G_c})$.
- Model 2 - reduced set of interactions: the J_1, J_2 and J_c genetic variants used to generate X_1 and X_2, plus the interactions from the product $\boldsymbol{G_1} \times \boldsymbol{G_2}$.

- Model 3 - continuous gene scores: the two weighted gene scores GS_{X_1} and GS_{X_2}, and their product $GS_{X_1} \times GS_{X_2}$.
- Model 4 - dichotomized gene scores: the two dichotomized gene scores, and their product.

Model 1 represents the 'gold standard' (oracle) model as it includes all of the variables used in the data generating model, whereas Models 2-4 are misspecified and their performance should be compared to Model 1. However, we do discuss the robustness of Model 1 to weak instrument bias. In Model 2, we have included a subset of the product terms between the genetic variants to create a more realistic scenario where the full set of relevant IVs are not included in the analysis. Model 3 considers the effect of including all of the genetic variants into two separate weighted gene scores, and finally, Model 4 considers the impact of dichotomizing the weighted gene scores as suggested by Ference *et al.* [33] when the authors were investigating interaction effects between drug treatments.

The data was generated 10 000 times with $\theta_0 = 0.2$, $\theta_1 = 0.3$, $\theta_2 = 0.2$, and $\theta_{12} = 0.1, 0.3$ and 0.5. A range of values for θ_{12} were chosen to consider the impact this had on the results, and to ensure they were comparable to θ_1 and θ_2. Each risk factor was associated with $(J_1 + J_c) = (J_2 + J_c) = 10$ genetic variants, and the number of cross-over variants J_c was initially set to 0 to consider the scenario where none of the genetic variants were associated with both risk factors. The correlation between the predicted values of the risk factors and the predicted values of the product term from the first stage regression were recorded. The analyses were re-performed on the mean centred risk factors with $\theta_{12} = 0.3$, and the number of cross-over variants set to $J_c = 1, 3, 5, 8$ and 10. A range of values for J_c were chosen to allow the performance of the models to be assessed on a sliding scale, starting with no genetic variants being associated with both risk factors, to all of the genetic variants being associated with both risk factors.

The following measurements were recorded for the estimates of θ_1, θ_2 and θ_{12}: median value; relative median bias; overall bias; standard deviation; median standard error; power at the 5% significance level; and coverage of the 95% confidence interval. The data was regenerated for $\sigma_1^2 = \sigma_2^2 = 5\%$ and 1%, when $J_c = 0, 5$ and 8, and the analyses were reperformed on the directly generated values of the risk factors, and estimates of the F-statistic and conditional F-statistic for X_1, X_2 and X_{12} were recorded.

Results The number of genetic variants used as IVs in Models 1 and 2 are displayed in Table 5.3. When none of the variants were associated with both risk factors ($J_c = 0$), Models 1 and 2 were equivalent and had 10 IVs for X_1, 10 IVs for X_2 and 10×10 IVs for X_{12} in the model. As the number of cross-over variants increased, the number of IVs in both models reduced, and this was particularly true for Model 2. When all of the variants were associated with both X_1 and X_2, only 10 IVs were included in Model 2, whereas Model 1 had 64 IVs. Although Models 3 and 4 used three IVs, both gene scores GS_{X_1} and GS_{X_2} contained information on 10 genetic variants irrespective of the value of J_c.

Table 5.3 Number of instrumental variables for Models 1 and 2 by the number of cross-over variants J_c.

	J_c $(J_1 = J_2)$[a]					
	0 (10)	1 (9)	3 (7)	5 (5)	8 (2)	10 (0)
Model 1 - full set of interactions						
Total number	120	119	114	105	84	65
First order	20	19	17	15	12	10
Interactions and quadratics	100	100	97	90	72	55
Model 2 - reduced set of interactions						
Total number	120	100	66	40	16	10
First order	20	19	17	15	12	10
Interactions	100	81	49	25	4	0

[a]Where J_c represents the number of genetic variants that were associated with both X_1 and X_2, and J_1 and J_2 represent the number of genetic variants associated with X_1 and X_2 respectively.

The results from the simulation study using 10 000 simulated datasets when $J_c = 0$ are presented in Table 5.4 (directly generated values of the risk factors) and Table 5.5 (mean centred values of the risk factors). For each model, the median estimate, the relative median bias of the estimate, the bias of the estimate, the standard deviation of the estimate, the median standard error, the statistical power to detect the effect at a nominal 5% significance level, the coverage of the true effect for a 95% confidence interval, and the median correlation between the predicted values from the first stage regression were recoded in Tables 5.4 and 5.5. Figure 5.3 contains histograms of the estimates and standard errors of the interaction term from Models 1, 3 and 4, when $J_c = 0$ and $\theta_{12} = 0.3$.

As expected, mean centring the risk factors substantially reduced the correlation between the predicted values of the risk factors with the predicted values of the product term, which helped to improve the precision of the main effect estimates, but had

no impact on the interaction term. Although the coverage of the main effects in Table 5.4 were at the nominal 95% level, there was under coverage in the marginal effects, and this increased in severity as θ_{12} increased from 0.1 to 0.5 (Table 5.5) due to the increased variability in the marginal effects (Equations 5.10 and 5.11). The coverage of the interaction term was close to the nominal level of 95% for all models.

In terms of precision and relative median bias of the interaction term, Model 1/2 (when $J_c = 0$ Models 1 and 2 are equivalent) out performed Models 3 and 4 (Figure 5.3). When $\theta_{12} = 0.3$, Model 1/2 had 98.8% power to detect the interaction term, whereas Models 3 and 4 had 77.5% and 41.8% respectively. Models 3 and 4 produced consistent estimates of the marginal and interaction effects despite using the externally weighted gene scores as IVs. There were noticeable differences in the power to detect the interaction term when the gene scores were treated as binary variables (Model 4), rather than continuous variables (Model 3). Since the gene scores were generated under the 'best-case' scenario with the individual level data and weights being generated from the same model, the discrepancy between the power of the continuous and binary gene scores to detect the interaction term may increase under more realistic scenarios. When $\theta_{12} = 0.1$, Model 4 only had 8% power to detect the interaction effect, compared to 33.7% and 15.4% for Models 1/2 and 3 respectively. Although the standard errors of the interaction term for Model 1/2 were symmetric, the standard errors for Models 3 and 4 were skewed (Figure 5.3).

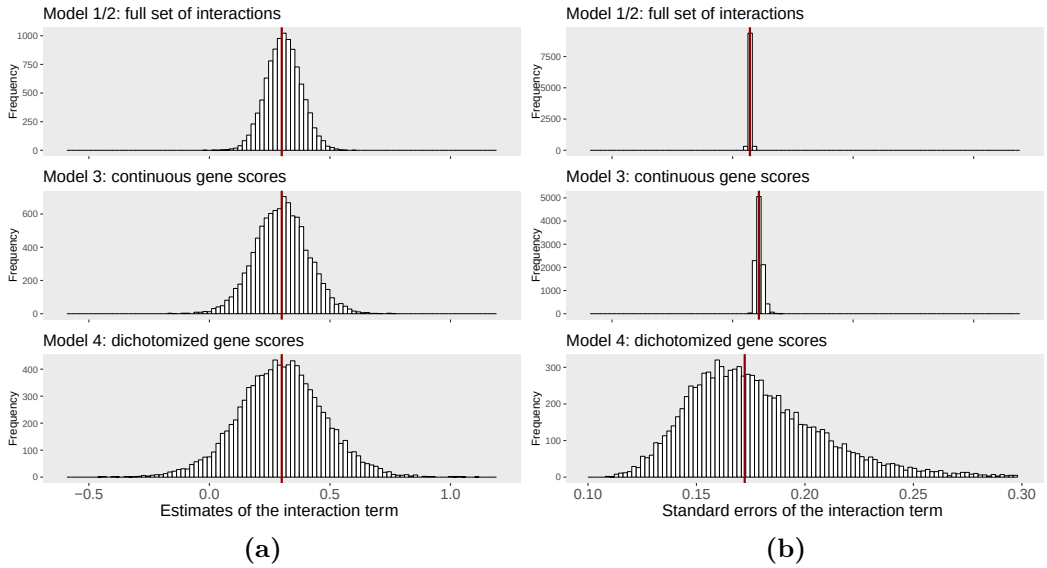


Fig. 5.3 Estimates of the interaction effect (a) and its standard error (b) from the two stage least squares regression models when there were no cross-over variants $J_c = 0$ and $\theta_{12} = 0.3$, and the mean centred values of the risk factors ($X_1 - \bar{X}_1$, $X_2 - \bar{X}_2$, $X_{12} = (X_1 - \bar{X}_1) \times (X_2 - \bar{X}_2)$) were used in the models. The red lines in (a) represent the true causal effect, and the red lines in (b) represent the median standard error.

Table 5.4 Performance of the two stage least squares regression models with respect to θ_1, θ_2 and θ_{12} when $J_c = 0$, and the directly generated values of the risk factors (X_1, X_2, $X_{12} = X_1 \times X_2$) were used in the models.

	med($\hat{\theta}$)	Rel bias	Bias	SD	SE	Power	Cov	Cor[a]
Models 1 & 2[b] - full set of interactions								
θ_1=0.3	0.3013	0.2009	0.0017	0.0917	0.0910	90.2	95.0	0.6694
θ_2=0.2	0.2022	0.3167	0.0019	0.0952	0.0945	57.1	94.9	0.6997
θ_{12}=**0.1**	0.1101	0.4882	0.0096	0.0721	0.0718	33.7	94.6	-
θ_1=0.3	0.3043	0.2027	0.0042	0.0918	0.0910	91.0	95.0	0.6695
θ_2=0.2	0.2034	0.3172	0.0032	0.0947	0.0945	57.9	95.5	0.6991
θ_{12}=**0.3**	0.3080	0.1629	0.0077	0.0722	0.0718	98.8	95.2	-
θ_1=0.3	0.3048	0.2008	0.0044	0.0911	0.0909	90.7	95.2	0.6701
θ_2=0.2	0.2050	0.3109	0.0043	0.0944	0.0945	58.4	95.2	0.6987
θ_{12}=**0.5**	0.5073	0.0942	0.0076	0.0715	0.0718	100.0	95.2	-
Model 3 - continuous gene scores								
θ_1=0.3	0.2993	0.2944	-0.0018	0.1362	0.1333	61.4	95.4	0.6802
θ_2=0.2	0.1991	0.4692	-0.0011	0.1415	0.1386	30.9	95.5	0.7115
θ_{12}=**0.1**	0.1010	0.7270	0.0011	0.1113	0.1091	15.4	95.5	-
θ_1=0.3	0.2998	0.2964	0.0012	0.1359	0.1332	61.9	95.6	0.6804
θ_2=0.2	0.2019	0.4610	0.0008	0.1405	0.1387	31.5	95.8	0.7105
θ_{12}=**0.3**	0.3000	0.2435	-0.0012	0.1106	0.1091	77.5	95.8	-
θ_1=0.3	0.3004	0.2956	0.0001	0.1352	0.1331	61.5	95.4	0.6810
θ_2=0.2	0.2008	0.4610	0.0005	0.1409	0.1385	30.7	95.6	0.7102
θ_{12}=**0.5**	0.4995	0.1461	-0.0002	0.1107	0.1092	98.7	95.6	-
Model 4 - dichotomized gene scores								
θ_1=0.3	0.2986	0.4609	-0.0021	0.2155	0.2072	31.0	95.7	0.6855
θ_2=0.2	0.1989	0.7167	-0.0010	0.2246	0.2168	15.0	96.2	0.7162
θ_{12}=**0.1**	0.1022	1.1399	0.0010	0.1786	0.1720	8.0	95.9	-
θ_1=0.3	0.3039	0.4601	0.0036	0.2145	0.2074	32.1	95.8	0.6840
θ_2=0.2	0.2047	0.7220	0.0031	0.2236	0.2164	15.2	96.2	0.7149
θ_{12}=**0.3**	0.2972	0.3851	-0.0032	0.1777	0.1722	41.8	96.0	-
θ_1=0.3	0.3010	0.4618	-0.0008	0.2148	0.2073	31.4	96.2	0.6857
θ_2=0.2	0.2002	0.7250	-0.0002	0.2233	0.2163	15.3	96.1	0.7147
θ_{12}=**0.5**	0.5002	0.2309	0.0003	0.1776	0.1718	80.7	96.1	-

Abbreviations: Rel bias, relative median bias; SD, standard deviation; SE, standard error; Cov, coverage; Cor, correlation.

[a]Median correlation between the predicted values of the risk factor X_1 or X_2 and the predicted values of X_{12} from the first stage regression of the TSLS model.

[b]When $J_c = 0$, Models 1 and 2 are equivalent.

Table 5.5 Performance of the two stage least squares regression models with respect to θ_{1M}, θ_{2M} and θ_{12} when $J_c = 0$, and the mean centred values of the risk factors ($X_1 - \bar{X}_1$, $X_2 - \bar{X}_2$, $X_{12} = (X_1 - \bar{X}_1) \times (X_2 - \bar{X}_2)$) were used in the models.

	med($\hat{\theta}$)	Rel bias	Bias	SD	SE	Power	Cov	Cor[a]
Models 1 & 2[b] - full set of interactions								
θ_{1M}=0.4176	0.4311	0.0574	0.0136	0.0327	0.0320	100.0	92.3	0.0153
θ_{2M}=0.3226	0.3370	0.0757	0.0143	0.0328	0.0320	100.0	91.9	0.0146
θ_{12}=**0.1**	0.1101	0.4882	0.0096	0.0721	0.0718	33.7	94.6	-
θ_{1M}=0.6527	0.6679	0.0454	0.0147	0.0408	0.0320	100.0	84.9	0.0142
θ_{2M}=0.5677	0.5823	0.0518	0.0144	0.0413	0.0320	100.0	85.0	0.0155
θ_{12}=**0.3**	0.3080	0.1629	0.0077	0.0722	0.0718	98.8	95.2	-
θ_{1M}=0.8879	0.9044	0.0418	0.0163	0.0527	0.0320	100.0	74.4	0.0141
θ_{2M}=0.8129	0.8290	0.0467	0.0165	0.0528	0.0320	100.0	73.9	0.0148
θ_{12}=**0.5**	0.5073	0.0942	0.0076	0.0715	0.0718	100.0	95.2	-
Model 3 - continuous gene scores								
θ_{1M}=0.4176	0.4178	0.0555	-0.0000	0.0348	0.0343	100.0	94.7	0.0000
θ_{2M}=0.3226	0.3234	0.0728	0.0008	0.0349	0.0343	100.0	94.4	0.0010
θ_{12}=**0.1**	0.1010	0.7270	0.0011	0.1113	0.1091	15.4	95.5	-
θ_{1M}=0.6527	0.6539	0.0442	0.0013	0.0424	0.0343	100.0	88.9	-0.0015
θ_{2M}=0.5677	0.5691	0.0506	0.0012	0.0431	0.0343	100.0	87.9	0.0011
θ_{12}=**0.3**	0.3000	0.2435	-0.0012	0.1106	0.1091	77.5	95.8	-
θ_{1M}=0.8879	0.8906	0.0410	0.0029	0.0539	0.0343	100.0	78.5	-0.0020
θ_{2M}=0.8129	0.8165	0.0459	0.0033	0.0543	0.0343	100.0	79.0	-0.0001
θ_{12}=**0.5**	0.4995	0.1461	-0.0002	0.1107	0.1092	98.7	95.6	-
Model 4 - dichotomized gene scores								
θ_{1M}=0.4176	0.4173	0.0705	-0.0006	0.0438	0.0435	100.0	95.1	0.0002
θ_{2M}=0.3226	0.3236	0.0929	0.0008	0.0438	0.0434	100.0	95.0	0.0027
θ_{12}=**0.1**	0.1022	1.1399	0.0010	0.1786	0.1720	8.0	95.9	-
θ_{1M}=0.6527	0.6538	0.0511	0.0012	0.0496	0.0435	100.0	91.5	-0.0013
θ_{2M}=0.5677	0.5687	0.0594	0.0010	0.0506	0.0435	100.0	91.0	0.0034
θ_{12}=**0.3**	0.2972	0.3851	-0.0032	0.1777	0.1722	41.8	96.0	-
θ_{1M}=0.8879	0.8913	0.0458	0.0029	0.0597	0.0435	100.0	84.5	
θ_{2M}=0.8129	0.8165	0.0502	0.0033	0.0603	0.0435	100.0	84.5	0.0011
θ_{12}=**0.5**	0.5002	0.2309	0.0003	0.1776	0.1718	80.7	96.1	-

Abbreviations: Rel bias, relative median bias; SD, standard deviation; SE, standard error; Cov, coverage; Cor, correlation.

[a]Median correlation between the predicted values of the centred risk factor $X_1 - \bar{X}_1$ or $X_2 - \bar{X}_2$ and the predicted values of X_{12} from the first stage regression of the TSLS model.

[b]When $J_c = 0$, Models 1 and 2 are equivalent.

The number of participants contained in each cell of the 2×2 contingency table when the gene scores were dichotomized at their median values for Model 4 are contained in Table 5.6. The results from the simulated datasets for when $\theta_{12} = 0.3$ and $J_c = 0, 1, 3, 5, 8$ and 10 are presented in Table 5.7. For each model, the median estimate, the relative median bias of the estimate, the bias of the estimate, the standard deviation of the estimate, the median standard error, the statistical power to detect the effect at a nominal 5% significance level, and the coverage of the true effect for a 95% confidence interval for the interaction term are presented in Table 5.7, along with the total number of IVs included in the model. Figure 5.4 contains the median standard error of the interaction effect across the four models for the different values of J_c.

All of the models produced consistent estimates of θ_{12}, and had good coverage properties for all values of J_c (Table 5.7). Increasing J_c had no impact on the estimates from Model 1, or the power to detect the interaction term (Figure 5.4), which never dropped below 98.8%. However, increasing the number of cross-over variants did reduce the power to detect θ_{12} in Model 2, where the power decreased from 96.7% ($J_c = 1$) to 8.7% ($J_c = 10$).

When the gene scores were treated as continuous variables (Model 3), the power to detect the interaction term increased from 77.5% ($J_c = 1$) to 93% ($J_c = 8$), and then dropped to 42.7% when $J_c = 10$. When the gene scores were dichotomized (Model 4), the power to detect θ_{12} decreased from 42.2% ($J_c = 1$) to 0.7% ($J_c = 10$). This reduction in power is reflected in Table 5.6, where the mean numbers of participants in groups n_{00}, n_{10}, n_{01}, and n_{11} were more unbalanced for larger values of J_c. When $J_c = 0$, the number of participants were evenly distributed across the four groups, but when $J_c = 10$, n_{10} and n_{01} only contained 2.2% of the data each.

Table 5.6 Mean numbers (%) of participants in the groups n_{00}, n_{10}, n_{01}, and n_{11} when the gene scores were dictohomized at their median values by the number of cross-over variants J_c.

J_c ($J_1 = J_2$)[a]	Mean numbers (%) of participants			
	n_{00}	n_{10}	n_{01}	n_{11}
0 (10)	2,502.3 (25.0)	2,497.7 (25.0)	2,497.7 (25.0)	2,502.3 (25.0)
1 (9)	2,658.0 (26.6)	2,342.0 (23.4)	2,342.0 (23.4)	2,658.0 (26.6)
3 (7)	2,979.9 (29.8)	2,020.1 (20.2)	2,020.1 (20.2)	2,979.9 (29.8)
5 (5)	3,323.0 (33.2)	1,677.0 (16.8)	1,677.0 (16.8)	3,322.9 (33.2)
8 (2)	3,951.9 (39.5)	1,048.5 (10.5)	1,048.5 (10.5)	3,951.2 (39.5)
10 (0)	4,781.6 (47.8)	220.1 (2.2)	220.1 (2.2)	4,778.3 (47.8)

[a]Where J_c represents the number of genetic variants that were associated with both X_1 and X_2, and J_1 and J_2 represent the number of genetic variants associated with X_1 and X_2 respectively.

Table 5.7 Performance of the four two stage least squares regression models with respect to the interaction when $\theta_{12} = 0.3$ and $J_c = 0, 1, 3, 5, 8$ and 10.

med($\hat{\theta}_{12}$)	Rel bias	Bias	SD	SE	Power	Coverage	J_c	IV total
Model 1 - full set of interactions								
0.3080	0.1629	0.0077	0.0722	0.0718	98.8	95.2	0[a]	120
0.3080	0.1647	0.0081	0.0723	0.0719	98.8	95.0	1	119
0.3090	0.1654	0.0081	0.0717	0.0716	98.9	95.3	3	114
0.3078	0.1655	0.0077	0.0716	0.0707	98.9	94.9	5	105
0.3073	0.1548	0.0071	0.0682	0.0687	99.3	95.2	8	84
0.3056	0.1497	0.0055	0.0670	0.0673	99.2	95.3	10	65
Model 2 - reduced set of interactions								
0.3073	0.1800	0.0074	0.0804	0.0794	96.7	94.9	1	100
0.3088	0.2225	0.0077	0.1003	0.0997	86.1	95.2	3	66
0.3056	0.2958	0.0058	0.1340	0.1334	63.2	95.7	5	40
0.3054	0.5389	0.0068	0.2520	0.2471	23.9	97.1	8	16
0.3057	0.7879	0.0091	0.3883	0.3891	8.7	99.3	10	10
Model 3 - continuous gene scores								
0.3000	0.2435	-0.0012	0.1106	0.1091	77.5	95.8	0	3
0.3005	0.2446	-0.0002	0.1111	0.1088	77.8	95.4	1	3
0.2998	0.2305	-0.0010	0.1051	0.1048	81.0	95.6	3	3
0.3015	0.2199	0.0019	0.0997	0.0980	85.6	95.5	5	3
0.3003	0.1889	0.0005	0.0857	0.0858	93.0	95.8	8	3
0.2993	0.2854	-0.3764	32.31	0.1711	42.7	99.2	10	3
Model 4 - dichotomized gene scores								
0.2972	0.3851	-0.0032	0.1777	0.1722	41.8	96.0	0	3
0.3028	0.3780	-0.0002	0.1757	0.1724	42.2	96.3	1	3
0.3002	0.3874	0.0005	0.1818	0.1773	39.8	96.4	3	3
0.3005	0.4177	0.0016	0.1948	0.1884	36.6	96.6	5	3
0.3007	0.5236	0.0009	0.2474	0.2340	25.7	97.2	8	3
0.2896	1.855	0.5258	133.5	1.3578	0.7	100.0	10	3

Abbreviations: Rel bias, relative median bias; SD, standard deviation; SE, standard error; Cov, coverage; IV, instrumental variable.
[a]When $J_c = 0$, Models 1 and 2 are equivalent.

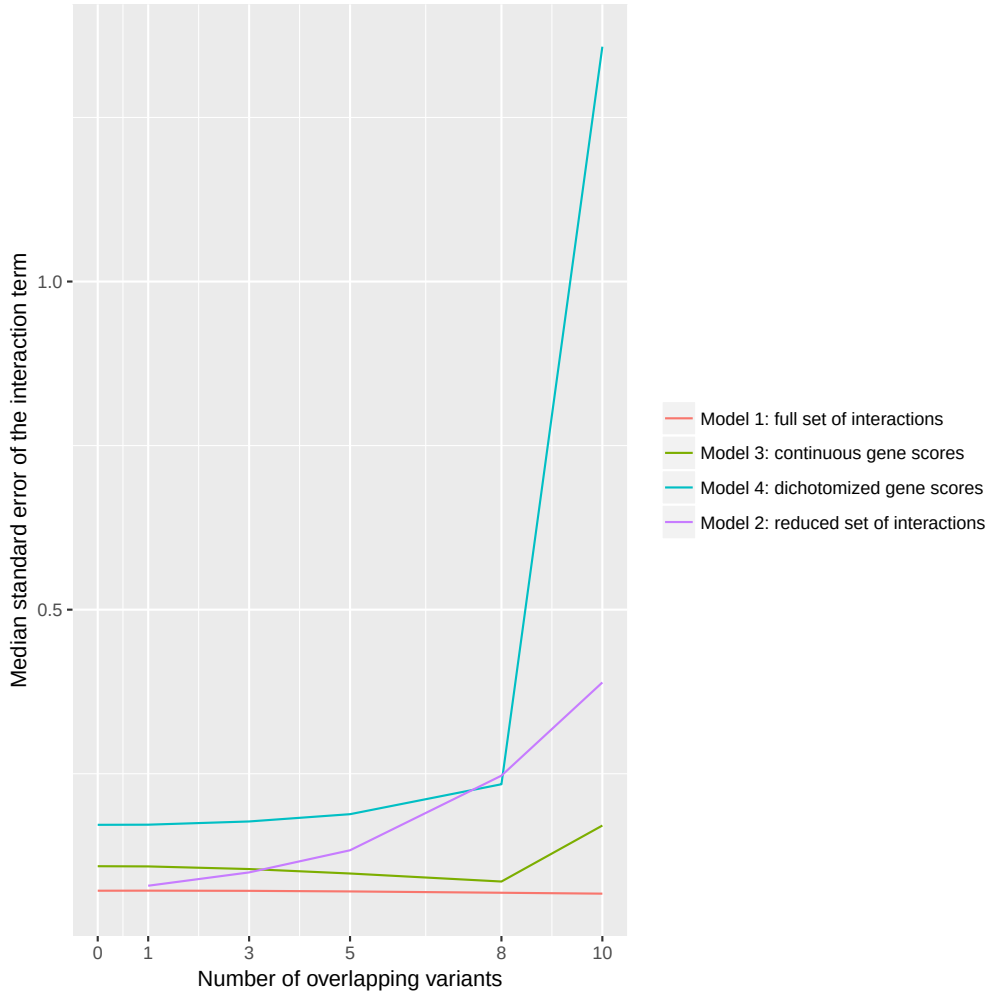


Fig. 5.4 Median standard errors of the interaction term when $\theta_{12} = 0.3$ for the four two stage least squares regression models when J_c was 0, 1, 3, 5, 8 and 10.

The results from the simulated datasets for when $\theta_1 = 0.3$, and the amount of variance in X_1 and X_2 explained by the genetic variants was σ_1^2=σ_2^2=10%, 5% and 1% are displayed in Table 5.8 ($J_c = 0$), Table 5.9 ($J_c = 5$), and Table 5.10 ($J_c = 8$). For θ_1, θ_2 and θ_{12}, the median estimate, the relative median bias of the estimate, the bias of the estimate, the standard deviation of the estimate, the median standard error, the statistical power to detect the effect at a nominal 5% significance level, and the coverage of the true effect for a 95% confidence interval are presented in Tables 5.8 to 5.10. The mean F-statistic and mean Sanderson-Windmeijer conditional F-statistic for X_1, X_2 and X_{12} are also recorded in Tables 5.8 to 5.10. For ease of comparison, the mean F-statistic, mean conditional F-statistic, and median estimates for θ_1, θ_2 and θ_{12} are presented in Table 5.11 for all values of σ_1^2, σ_2^2 and J_c.

As the number of cross-over variants increased, the mean F-statistic increased for all models when σ_1^2=σ_2^2=10%, 5% and 1% (Table 5.11). Except for Model 4, the conditional F-statistic either increased or remained the same with more cross-over variants. Increasing J_c had a greater influence on the conditional F-statistic for X_{12} than X_1 or X_2. For all scenarios considered, Model 1 consistently had the smallest mean F-statistic and mean conditional F-statistic, whilst Model 3 always had the largest.

As the amount of variance explained by the genetic variants in X_1 and X_2 decreased from 10% to 1%, the mean F-statistic and conditional F-statistic decreased for all models when $J_c = 0, 5$ and 8. When the mean F-statistic was less than 10, the main effect terms θ_1 and θ_2 suffered from weak instrument bias for Models 1 and 2. The mean F-statistic never dropped below 10 for Models 3 and 4 for all of the scenarios considered. The reduction in the mean F-statistic as σ_1^2 and σ_2^2 decreased had no impact on the estimates for θ_{12} for Models 1 or 2. There was slight bias in the estimates for θ_{12} when $\sigma_1^2 = \sigma_2^2$=1% for Models 3 and 4.

As the number of cross-over variants increased, the power to detect the main effects θ_1 and θ_2 decreased for all models except Model 3 (Tables 5.8 to 5.10). The impact of increasing J_c on the power to detect θ_{12} across the four TSLS regression models has already been discussed (Table 5.7), and the same conclusions are observed for σ_1^2=σ_2^2=5% and 1%. As σ_1^2 and σ_2^2 decreased, the power to detect the interaction term decreased for all models. Model 1 was least affected by the reduction in variance explained, whereas the power to detect θ_{12} in Models 3 and 4 decreased significantly for all values of J_c as σ_1^2 and σ_2^2 decreased. When $J_c = 0$, the power to detect the interaction term for Model 3 when the variants explained 10% of the variance in the risk factors was 77.5%, and this reduced to 0.7% when 1% of the variance was explained. The lowest power to detect the interaction for Model 1 was 74.7% when $J_c = 10$, and 1% of the variances in X_1 and X_2 were explained.

Table 5.8 Performance of the two stage least squares regression models by the amount of variance the genetic variants $\boldsymbol{G_1}$ and $\boldsymbol{G_2}$ explained in X_1 and X_2 when $\theta_1 = 0.3$, $\theta_2 = 0.2$, $\theta_{12} = 0.3$ and $J_c = 0$, and the directly generated values of the risk factors (X_1, X_2, $X_{12} = X_1 \times X_2$) were used in the models.

	F-stat[a]	CF-stat[a]	med($\hat{\theta}$)	Rel bias	Bias	SD	SE	Pow	Cov
G_1 and G_2 explained 10% of the variance in X_1 and X_2									
Models 1 & 2[b] - full set of interactions									
θ_1	10.3 (0.6)	2.1 (0.3)	0.3043	0.2027	0.0042	0.0918	0.0910	91.0	95.0
θ_2	10.3 (0.6)	2.1 (0.3)	0.2034	0.3172	0.0032	0.0947	0.0945	57.9	95.5
θ_{12}	8.1 (0.6)	1.9 (0.2)	0.3080	0.1629	0.0077	0.0722	0.0718	98.8	95.2
Model 3 - continuous gene scores									
θ_1	364.2 (23.4)	104.5 (25.6)	0.2998	0.2964	0.0012	0.1359	0.1332	61.9	95.6
θ_2	364.5 (23.2)	103.9 (25.3)	0.2019	0.4610	0.0008	0.1405	0.1387	31.5	95.8
θ_{12}	273.7 (22.4)	97.8 (22.8)	0.3000	0.2435	-0.0012	0.1106	0.1091	77.5	95.8
Model 4 - dichotomized gene scores									
θ_1	224.2 (17.7)	41.9 (13.4)	0.3039	0.4601	0.0036	0.2145	0.2074	32.1	95.8
θ_2	224.4 (17.7)	41.7 (13.3)	0.2047	0.7220	0.0031	0.2236	0.2164	15.2	96.2
θ_{12}	168.2 (16.3)	40.0 (12.4)	0.2972	0.3851	-0.0032	0.1777	0.1722	41.8	96.0
G_1 and G_2 explained 5% of the variance in X_1 and X_2									
Models 1 & 2[b] - full set of interactions									
θ_1	5.4 (0.4)	1.5 (0.2)	0.3174	0.2093	0.0169	0.0931	0.0920	92.4	94.5
θ_2	5.4 (0.4)	1.4 (0.2)	0.2166	0.3220	0.0169	0.0957	0.0959	62.0	94.8
θ_{12}	3.9 (0.4)	1.2 (0.2)	0.3087	0.1981	0.0074	0.0889	0.0888	92.8	95.0
Model 3 - continuous gene scores									
θ_1	170.2 (15.5)	25.4 (11.7)	0.2988	0.4733	-0.0026	0.2298	0.2121	29.9	96.9
θ_2	170.1 (15.7)	25.2 (11.5)	0.1985	0.7494	-0.0026	0.2421	0.2237	13.8	96.9
θ_{12}	109.4 (13.3)	23.8 (10.4)	0.3020	0.5061	0.0020	0.2458	0.2276	26.7	96.9
Model 4 - dichotomized gene scores									
θ_1	107.3 (12.2)	10.7 (6.7)	0.2970	0.7506	0.0382	3.928	0.3367	12.6	98.9
θ_2	106.9 (12.0)	10.6 (6.6)	0.1948	1.181	0.0357	3.804	0.3551	5.4	98.7
θ_{12}	68.8 (10.2)	10.2 (6.1)	0.3033	0.8104	-0.0404	4.065	0.3654	10.8	98.8
G_1 and G_2 explained 1% of the variance in X_1 and X_2									
Models 1 & 2[b] - full set of interactions									
θ_1	1.8 (0.2)	1.4 (0.2)	0.3681	0.2637	0.0677	0.0910	0.0901	97.7	88.4
θ_2	1.8 (0.2)	1.4 (0.2)	0.2670	0.3987	0.0673	0.0930	0.0930	81.4	88.6
θ_{12}	1.4 (0.2)	1.0 (0.1)	0.3029	0.2152	0.0012	0.0971	0.0972	86.4	95.4
Model 3 - continuous gene scores									
θ_1	29.5 (6.4)	1.9 (2.9)	0.2854	1.446	0.0373	29.26	0.8411	2.8	99.9
θ_2	29.4 (6.4)	1.9 (2.8)	0.1883	2.379	0.0320	31.58	0.9203	1.0	99.9
θ_{12}	12.3 (4.1)	1.6 (2.1)	0.3185	2.655	-0.0196	52.32	1.537	0.7	100.0
Model 4 - dichotomized gene scores									
θ_1	19.1 (5.1)	1.6 (2.8)	0.2992	1.690	-0.0787	123.8	1.063	1.9	99.9
θ_2	19.0 (5.0)	1.5 (2.4)	0.1930	2.785	0.7563	217.5	1.163	0.6	100.0
θ_{12}	8.1 (3.3)	1.2 (1.7)	0.3121	3.069	-1.376	347.4	1.933	0.3	100.0

Abbreviations: F-stat, F-statistic; CF-stat, conditional F-statistic; Rel bias, relative median bias; SD, standard deviation; SE, standard error; Pow, power; Cov, coverage.

[a]The F-statistic and conditional F-statistic were calculated for X_1, X_2, and X_{12}.

[b]When $J_c = 0$, Models 1 and 2 are equivalent.

Table 5.9 Performance of the two stage least squares regression models by the amount of variance the genetic variants $\boldsymbol{G_1}$, $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$ explained in X_1 and X_2 when $\theta_1 = 0.3$, $\theta_2 = 0.2$, $\theta_{12} = 0.3$ and $J_c = 5$, and the directly generated values of the risk factors (X_1, X_2, $X_{12} = X_1 \times X_2$) were used in the models.

	F-stat[a]	CF-stat[a]	med($\hat{\theta}$)	Rel bias	Bias	SD	SE	Pow	Cov
$\boldsymbol{G_1} + \boldsymbol{G_c}$ and $\boldsymbol{G_2} + \boldsymbol{G_c}$ explained 10% of the variance in X_1 and X_2									
Model 1 - full set of interactions									
θ_1	11.6 (0.7)	2.5 (0.4)	0.2981	0.2091	-0.0011	0.0933	0.0927	89.1	95.0
θ_2	11.6 (0.7)	2.5 (0.4)	0.1988	0.3178	-0.0015	0.0955	0.0960	55.0	95.5
θ_{12}	13.4 (0.9)	2.2 (0.3)	0.3074	0.1588	0.0076	0.0707	0.0706	99.0	95.0
Model 2 - reduced set of interactions									
θ_1	28.8 (1.8)	2.6 (0.4)	0.2970	0.3677	-0.0045	0.1664	0.1649	44.2	95.8
θ_2	28.8 (1.8)	2.6 (0.4)	0.1966	0.5672	-0.0049	0.1719	0.1715	21.0	95.9
θ_{12}	32.6 (2.1)	2.3 (0.3)	0.3056	0.2965	0.0062	0.1337	0.1333	63.4	95.8
Model 3 - continuous gene scores									
θ_1	366.4 (23.2)	131.8 (30.9)	0.2993	0.2848	-0.0006	0.1272	0.1244	67.0	95.4
θ_2	366.3 (23.4)	131.0 (30.7)	0.1992	0.4385	-0.0005	0.1314	0.1293	35.1	95.4
θ_{12}	426.6 (29.1)	120.9 (26.9)	0.3008	0.2240	0.0003	0.1000	0.0978	84.8	95.4
Model 4 - dichotomized gene scores									
θ_1	233.5 (18.1)	35.8 (12.4)	0.2984	0.5087	-0.0027	0.2399	0.2302	25.9	96.4
θ_2	233.5 (18.2)	35.6 (12.3)	0.2005	0.7954	-0.0019	0.2482	0.2396	13.0	96.4
θ_{12}	284.1 (21.6)	33.8 (11.2)	0.3006	0.4126	0.0017	0.1950	0.1877	36.8	96.4
$\boldsymbol{G_1} + \boldsymbol{G_c}$ and $\boldsymbol{G_2} + \boldsymbol{G_c}$ explained 5% of the variance in X_1 and X_2									
Model 1 - full set of interactions									
θ_1	6.0 (0.5)	1.6 (0.2)	0.3052	0.2184	0.0073	0.0980	0.0983	87.7	95.2
θ_2	6.0 (0.5)	1.5 (0.2)	0.2078	0.3347	0.0069	0.1018	0.1022	53.3	95.2
θ_{12}	6.1 (0.5)	1.3 (0.2)	0.3097	0.2065	0.0088	0.0925	0.0919	91.3	95.3
Model 2 - reduced set of interactions									
θ_1	14.2 (1.2)	1.6 (0.3)	0.2982	0.3577	-0.0005	0.1600	0.1588	48.4	96.3
θ_2	14.2 (1.2)	1.6 (0.3)	0.1994	0.5431	-0.0001	0.1665	0.1664	22.7	96.1
θ_{12}	13.9 (1.3)	1.4 (0.2)	0.3087	0.3591	0.0071	0.1621	0.1615	49.0	96.1
Model 3 - continuous gene scores									
θ_1	171.8 (15.6)	32.9 (14.1)	0.3014	0.4376	0.0012	0.2078	0.1951	35.7	96.4
θ_2	172.1 (15.4)	32.6 (13.9)	0.2041	0.6828	0.0026	0.2169	0.2043	16.9	96.5
θ_{12}	171.7 (17.6)	30.0 (12.0)	0.2981	0.4558	-0.0021	0.2147	0.2010	32.6	96.5
Model 4 - dichotomized gene scores									
θ_1	111.9 (12.5)	9.5 (6.4)	0.2933	0.8209	-0.0136	0.8024	0.3732	10.2	99.1
θ_2	112.2 (12.3)	9.4 (6.3)	0.1981	1.2850	-0.0120	0.8127	0.3926	4.6	98.9
θ_{12}	117.6 (13.5)	8.9 (5.7)	0.3066	0.8749	0.0135	0.8619	0.3967	9.6	99.1
$\boldsymbol{G_1} + \boldsymbol{G_c}$ and $\boldsymbol{G_2} + \boldsymbol{G_c}$ explained 1% of the variance in X_1 and X_2									
Model 1 - full set of interactions									
θ_1	2.0 (0.2)	1.4 (0.2)	0.3504	0.2507	0.0502	0.0975	0.0971	94.4	92.0
θ_2	2.0 (0.2)	1.3 (0.2)	0.2478	0.3740	0.0485	0.1003	0.1002	69.6	92.2
θ_{12}	1.6 (0.2)	1.0 (0.1)	0.3037	0.2362	0.0028	0.1051	0.1043	82.0	95.3
Model 2 - reduced set of interactions									
θ_1	3.5 (0.6)	1.4 (0.2)	0.3225	0.3105	0.0223	0.1398	0.1395	63.8	95.6
θ_2	3.5 (0.5)	1.4 (0.2)	0.2243	0.4832	0.0227	0.1459	0.1457	34.3	95.7
θ_{12}	2.6 (0.5)	1.1 (0.1)	0.3036	0.3886	0.0040	0.1771	0.1758	41.8	96.1
Model 3 - continuous gene scores									
θ_1	31.0 (6.6)	2.5 (3.7)	0.2912	1.353	0.4689	47.33	0.7448	3.6	99.9
θ_2	30.9 (6.5)	2.3 (3.4)	0.1939	2.201	0.4324	41.15	0.8014	1.1	99.9
θ_{12}	19.9 (5.4)	1.9 (2.4)	0.3030	2.399	-0.6582	72.69	1.315	0.6	99.9
Model 4 - dichotomized gene scores									
θ_1	20.9 (5.3)	1.6 (2.9)	0.2967	1.760	0.4713	65.97	1.108	1.5	99.9
θ_2	20.8 (5.2)	1.5 (2.5)	0.1959	2.887	0.4772	54.84	1.208	0.4	100.0
θ_{12}	14.1 (4.4)	1.2 (1.6)	0.3096	3.205	-0.8906	105.7	1.991	0.2	100.0

Abbreviations: F-stat, F-statistic; CF-stat, conditional F-statistic; Rel bias, relative median bias; SD, standard deviation; SE, standard error; Pow, power; Cov, coverage.

[a]The F-statistic and conditional F-statistic were calculated for X_1, X_2, and X_{12}.

Table 5.10 Performance of the two stage least squares regression models by the amount of variance the genetic variants $\boldsymbol{G_1}$, $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$ explained in X_1 and X_2 when $\theta_1 = 0.3$, $\theta_2 = 0.2$, $\theta_{12} = 0.3$ and $J_c = 8$, and the directly generated values of the risk factors (X_1, X_2, $X_{12} = X_1 \times X_2$) were used in the models.

	F-stat[a]	CF-stat[a]	med($\hat{\theta}$)	Rel bias	Bias	SD	SE	Pow	Cov
$\boldsymbol{G_1} + \boldsymbol{G_c}$ and $\boldsymbol{G_2} + \boldsymbol{G_c}$ explained 10% of the variance in X_1 and X_2									
Model 1 - full set of interactions									
θ_1	14.2 (0.9)	3.2 (0.4)	0.2966	0.2143	-0.0029	0.0965	0.0967	86.1	95.4
θ_2	14.2 (0.8)	3.2 (0.4)	0.1971	0.3324	-0.0038	0.0993	0.0997	50.9	95.4
θ_{12}	19.9 (1.2)	3.0 (0.4)	0.3073	0.1548	0.0071	0.0682	0.0687	99.3	95.2
Model 2 - reduced set of interactions									
θ_1	70.4 (4.4)	3.6 (0.5)	0.2949	0.6673	-0.0071	0.3126	0.3047	16.1	97.0
θ_2	70.5 (4.4)	3.6 (0.5)	0.1914	1.0321	-0.0075	0.3233	0.3168	8.3	97.3
θ_{12}	96.2 (5.8)	3.4 (0.5)	0.3054	0.5389	0.0068	0.2520	0.2471	23.9	97.1
Model 3 - continuous gene scores									
θ_1	367.3 (23.5)	168.0 (33.0)	0.2989	0.2616	-0.0004	0.1190	0.1191	71.4	95.5
θ_2	368.0 (23.2)	167.3 (32.8)	0.1996	0.4042	-0.0011	0.1223	0.1227	36.9	95.3
θ_{12}	520.7 (32.8)	156.4 (31.7)	0.3003	0.1889	0.0005	0.0857	0.0858	93.0	95.8
Model 4 - dichotomized gene scores									
θ_1	243.2 (18.7)	24.2 (10.3)	0.2979	0.6628	-0.0023	0.3134	0.2981	16.4	97.0
θ_2	243.8 (18.7)	24.1 (10.1)	0.2007	1.018	-0.0008	0.3253	0.3091	8.5	97.0
θ_{12}	340.8 (24.2)	22.1 (8.9)	0.3007	0.5236	0.0009	0.2474	0.2340	25.7	97.2
$\boldsymbol{G_1} + \boldsymbol{G_c}$ and $\boldsymbol{G_2} + \boldsymbol{G_c}$ explained 5% of the variance in X_1 and X_2									
Models 1 - full set of interactions									
θ_1	7.3 (0.6)	1.7 (0.3)	0.3047	0.2430	0.0035	0.1098	0.1095	79.2	95.2
θ_2	7.3 (0.6)	1.7 (0.3)	0.2039	0.3747	0.0041	0.1122	0.1135	43.7	95.6
θ_{12}	8.7 (0.7)	1.5 (0.2)	0.3072	0.2139	0.0070	0.0962	0.0966	88.6	95.4
Model 2 - reduced set of interactions									
θ_1	33.9 (3.0)	2.0 (0.3)	0.2977	0.5994	-0.0011	0.2825	0.2759	18.6	97.8
θ_2	33.9 (3.0)	1.9 (0.3)	0.1974	0.9480	0.0012	0.2964	0.2896	8.9	97.8
θ_{12}	39.7 (3.5)	1.7 (0.3)	0.3028	0.6108	0.0022	0.2906	0.2849	18.1	98.1
Model 3 - continuous gene scores									
θ_1	173.1 (15.7)	44.6 (16.9)	0.2998	0.4113	0.0003	0.1926	0.1844	38.0	96.1
θ_2	173.1 (15.6)	44.0 (16.7)	0.2035	0.6260	0.0026	0.1973	0.1920	18.2	96.2
θ_{12}	209.5 (19.7)	38.5 (14.1)	0.2977	0.3901	-0.0016	0.1832	0.1759	40.5	96.5
Model 4 - dichotomized gene scores									
θ_1	117.5 (12.7)	6.7 (5.5)	0.3056	1.074	-0.2320	20.77	0.4940	5.7	99.3
θ_2	117.4 (12.8)	6.7 (5.4)	0.2061	1.664	-0.2843	25.93	0.5204	2.5	99.5
θ_{12}	142.4 (14.9)	5.9 (4.5)	0.2983	1.109	0.2829	25.47	0.5114	4.8	99.6
$\boldsymbol{G_1} + \boldsymbol{G_c}$ and $\boldsymbol{G_2} + \boldsymbol{G_c}$ explained 1% of the variance in X_1 and X_2									
Models 1 - full set of interactions									
θ_1	2.2 (0.3)	1.3 (0.2)	0.3368	0.2633	0.0374	0.1117	0.1105	85.8	94.0
θ_2	2.2 (0.3)	1.2 (0.2)	0.2362	0.3949	0.0372	0.1134	0.1140	54.8	94.0
θ_{12}	1.9 (0.3)	1.0 (0.2)	0.3056	0.2546	0.0050	0.1173	0.1169	74.7	95.2
Model 2 - reduced set of interactions									
θ_1	7.3 (1.3)	1.4 (0.2)	0.3113	0.4897	0.0087	0.2292	0.2261	28.5	97.4
θ_2	7.3 (1.3)	1.4 (0.2)	0.2077	0.7600	0.0074	0.2388	0.2373	12.7	97.6
θ_{12}	5.6 (1.2)	1.2 (0.2)	0.3066	0.6499	0.0033	0.3054	0.3042	15.6	98.2
Model 3 - continuous gene scores									
θ_1	31.9 (6.6)	2.9 (3.9)	0.3048	1.339	-0.7866	95.56	0.7330	2.9	99.8
θ_2	32.0 (6.6)	2.8 (3.8)	0.2129	2.117	-1.464	141.5	0.7834	1.0	99.8
θ_{12}	24.5 (5.9)	2.1 (2.5)	0.2868	2.170	2.167	220.6	1.196	0.8	100.0
Model 4 - dichotomized gene scores									
θ_1	22.5 (5.5)	1.3 (2.1)	0.3099	2.113	-0.4121	26.77	1.486	0.8	99.9
θ_2	22.5 (5.5)	1.3 (2.2)	0.2035	3.358	0.0140	27.80	1.619	0.2	100.0
θ_{12}	17.5 (4.8)	1.0 (1.4)	0.2851	3.203	0.4207	45.99	2.334	0.1	100.0

Abbreviations: F-stat, F-statistic; CF-stat, conditional F-statistic; Rel bias, relative median bias; SD, standard deviation; SE, standard error; Pow, power; Cov, coverage.

[a]The F-statistic and conditional F-statistic were calculated for X_1, X_2, and X_{12}.

Table 5.11 Summary of the performance of the two stage least squares regression models by the amount of variance the genetic variants $\boldsymbol{G_1}$, $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$ explained in X_1 and X_2 when $\theta_1 = 0.3$, $\theta_2 = 0.2$, $\theta_{12} = 0.3$ and $J_c = 0, 5$ and 8, and the directly generated values of the risk factors (X_1, X_2, $X_{12} = X_1 \times X_2$) were used in the models.

	10% variance explained			5% variance explained			1% variance explained		
	F-stat[a]	CF-stat[a]	med($\hat{\theta}$)	F-stat	CF-stat	med($\hat{\theta}$)	F-stat	CF-stat	med($\hat{\theta}$)
$J_c = 0$									
Models 1 & 2[b] - full set of interactions									
θ_1	10.3 (0.6)	2.1 (0.3)	0.3043	5.4 (0.4)	1.5 (0.2)	0.3174	1.8 (0.2)	1.4 (0.2)	0.3681
θ_2	10.3 (0.6)	2.1 (0.3)	0.2034	5.4 (0.4)	1.4 (0.2)	0.2166	1.8 (0.2)	1.4 (0.2)	0.2670
θ_{12}	8.1 (0.6)	1.9 (0.2)	0.3080	3.9 (0.4)	1.2 (0.2)	0.3087	1.4 (0.2)	1.0 (0.1)	0.3029
Model 3 - continuous gene scores									
θ_1	364.2 (23.4)	104.5 (25.6)	0.2998	170.2 (15.5)	25.4 (11.7)	0.2988	29.5 (6.4)	1.9 (2.9)	0.2854
θ_2	364.5 (23.2)	103.9 (25.3)	0.2019	170.1 (15.7)	25.2 (11.5)	0.1985	29.4 (6.4)	1.9 (2.8)	0.1883
θ_{12}	273.7 (22.4)	97.8 (22.8)	0.3000	109.4 (13.3)	23.8 (10.4)	0.3020	12.3 (4.1)	1.6 (2.1)	0.3185
Model 4 - dichotomized gene scores									
θ_1	224.2 (17.7)	41.9 (13.4)	0.3039	107.3 (12.2)	10.7 (6.7)	0.2970	19.1 (5.1)	1.6 (2.8)	0.2992
θ_2	224.4 (17.7)	41.7 (13.3)	0.2047	106.9 (12.0)	10.6 (6.6)	0.1948	19.0 (5.0)	1.5 (2.4)	0.1930
θ_{12}	168.2 (16.3)	40.0 (12.4)	0.2972	68.8 (10.2)	10.2 (6.1)	0.3033	8.1 (3.3)	1.2 (1.7)	0.3121
$J_c = 5$									
Models 1 - full set of interactions									
θ_1	11.6 (0.7)	2.5 (0.4)	0.2981	6.0 (0.5)	1.6 (0.2)	0.3052	2.0 (0.2)	1.4 (0.2)	0.3504
θ_2	11.6 (0.7)	2.5 (0.4)	0.1988	6.0 (0.5)	1.5 (0.2)	0.2078	2.0 (0.2)	1.3 (0.2)	0.2478
θ_{12}	13.4 (0.9)	2.2 (0.3)	0.3074	6.1 (0.5)	1.3 (0.2)	0.3097	1.6 (0.2)	1.0 (0.1)	0.3037
Model 2 - reduced set of interactions									
θ_1	28.8 (1.8)	2.6 (0.4)	0.2970	14.2 (1.2)	1.6 (0.3)	0.2982	3.5 (0.6)	1.4 (0.2)	0.3225
θ_2	28.8 (1.8)	2.6 (0.4)	0.1966	14.2 (1.2)	1.6 (0.3)	0.1994	3.5 (0.5)	1.4 (0.2)	0.2243
θ_{12}	32.6 (2.1)	2.3 (0.3)	0.3056	13.9 (1.3)	1.4 (0.2)	0.3087	2.6 (0.5)	1.1 (0.1)	0.3036
Model 3 - continuous gene scores									
θ_1	366.4 (23.2)	131.8 (30.9)	0.2993	171.8 (15.6)	32.9 (14.1)	0.3014	31.0 (6.6)	2.5 (3.7)	0.2912
θ_2	366.3 (23.4)	131.0 (30.7)	0.1992	172.1 (15.4)	32.6 (13.9)	0.2041	30.9 (6.5)	2.3 (3.4)	0.1939
θ_{12}	426.6 (29.1)	120.9 (26.9)	0.3008	171.7 (17.6)	30.0 (12.0)	0.2981	19.9 (5.4)	1.9 (2.4)	0.3030
Model 4 - dichotomized gene scores									
θ_1	233.5 (18.1)	35.8 (12.4)	0.2984	111.9 (12.5)	9.5 (6.4)	0.2933	20.9 (5.3)	1.6 (2.9)	0.2967
θ_2	233.5 (18.2)	35.6 (12.3)	0.2005	112.2 (12.3)	9.4 (6.3)	0.1981	20.8 (5.2)	1.5 (2.5)	0.1959
θ_{12}	284.1 (21.6)	33.8 (11.2)	0.3006	117.6 (13.5)	8.9 (5.7)	0.3066	14.1 (4.4)	1.2 (1.6)	0.3096
$J_c = 8$									
Models 1 - full set of interactions									
θ_1	14.2 (0.9)	3.2 (0.4)	0.2966	7.3 (0.6)	1.7 (0.3)	0.3047	2.2 (0.3)	1.3 (0.2)	0.3368
θ_2	14.2 (0.8)	3.2 (0.4)	0.1971	7.3 (0.6)	1.7 (0.3)	0.2039	2.2 (0.3)	1.2 (0.2)	0.2362
θ_{12}	19.9 (1.2)	3.0 (0.4)	0.3073	8.7 (0.7)	1.5 (0.2)	0.3072	1.9 (0.3)	1.0 (0.2)	0.3056
Model 2 - reduced set of interactions									
θ_1	70.4 (4.4)	3.6 (0.5)	0.2949	33.9 (3.0)	2.0 (0.3)	0.2977	7.3 (1.3)	1.4 (0.2)	0.3113
θ_2	70.5 (4.4)	3.6 (0.5)	0.1914	33.9 (3.0)	1.9 (0.3)	0.1974	7.3 (1.3)	1.4 (0.2)	0.2077
θ_{12}	96.2 (5.8)	3.4 (0.5)	0.3054	39.7 (3.5)	1.7 (0.3)	0.3028	5.6 (1.2)	1.2 (0.2)	0.3066
Model 3 - continuous gene scores									
θ_1	367.3 (23.5)	168.0 (33.0)	0.2989	173.1 (15.7)	44.6 (16.9)	0.2998	31.9 (6.6)	2.9 (3.9)	0.3048
θ_2	368.0 (23.2)	167.3 (32.8)	0.1996	173.1 (15.6)	44.0 (16.7)	0.2035	32.0 (6.6)	2.8 (3.8)	0.2129
θ_{12}	520.7 (32.8)	156.4 (31.7)	0.3003	209.5 (19.7)	38.5 (14.1)	0.2977	24.5 (5.9)	2.1 (2.5)	0.2868
Model 4 - dichotomized gene scores									
θ_1	243.2 (18.7)	24.2 (10.3)	0.2979	117.5 (12.7)	6.7 (5.5)	0.3056	22.5 (5.5)	1.3 (2.1)	0.3099
θ_2	243.8 (18.7)	24.1 (10.1)	0.2007	117.4 (12.8)	6.7 (5.4)	0.2061	22.5 (5.5)	1.3 (2.2)	0.2035
θ_{12}	340.8 (24.2)	22.1 (8.9)	0.3007	142.4 (14.9)	5.9 (4.5)	0.2983	17.5 (4.8)	1.0 (1.4)	0.2851

Abbreviations: F-stat, F-statistic; CF-stat, conditional F-statistic.
[a]The F-statistic and conditional F-statistic were calculated for X_1, X_2, and X_{12}.
[b]When $J_c = 0$, Models 1 and 2 are equivalent.

Summary level data

As data on genetic associations has become more publicly accessible, Mendelian randomization analyses tend to be implemented using summarized level data [142]. Summarized data on the associations of each genetic variant with the trait(s) may be obtained from a consortium: the beta-coefficients and their standard errors from univariable regression on each variant in turn. Typically, consortia do not provide summarized data on the product of two traits or the interaction effect of genetic variants with the trait(s). Using the same framework and notation outlined in Section 5.4.1, we show that it is not possible to extend the multivariable IVW model to estimate the interaction effect between two risk factors using standard summarized data (genetic associations of each variant with the risk factors and outcome).

Suppose we have summarized data on the genetic associations for the J genetic variants G_j with X_1, X_2, and Y, and we consider estimating θ_{12} in Equation 5.7 by including the product of the genetic associations of X_1 and X_2 in Equation 5.6:

$$\hat{\beta}_{Y_j} = \theta_1'\hat{\beta}_{X_1j} + \theta_2'\hat{\beta}_{X_2j} + \theta_{12}'\hat{\beta}_{X_1j} \times \hat{\beta}_{X_2j} + \epsilon_j\,, \quad \text{weights} = \text{se}(\hat{\beta}_{Y_j})^{-2}\,, \tag{5.12}$$

where $\hat{\beta}_{X_1j}$, $\hat{\beta}_{X_2j}$ and $\hat{\beta}_{Y_j}$ are estimates of the j genetic associations with X_1, X_2 and Y, and $\text{se}(\hat{\beta}_{Y_j})^{-2}$ are the standard errors of the genetic associations with Y. For simplicity, we assume that the MAF and the genetic associations are the same within each subgroup of the genetic variants ($\boldsymbol{G_1}$, $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$). From Section 5.4.1, the risk factors X_1 and X_2 can be expressed as:

$$\begin{aligned} X_1 &= \beta_{01} + \beta_1 \sum \boldsymbol{G_1} + \beta_{1c} \sum \boldsymbol{G_c} + U_1 + \epsilon_1 \quad \text{and} \\ X_2 &= \beta_{02} + \beta_2 \sum \boldsymbol{G_2} + \beta_{2c} \sum \boldsymbol{G_c} + U_2 + \epsilon_2\,, \end{aligned}$$

where we assume that the unmeasured confounders U_1 and U_2 are from a normal distribution with mean zero. If we let $\mu_1 = \beta_{01} + U_1$ and $\mu_2 = \beta_{02} + U_2$, then the product term X_{12} can be written as:

$$\begin{aligned} X_{12} = X_1 \times X_2 &= (\mu_1 + \beta_1 \sum \boldsymbol{G_1} + \beta_{1c} \sum \boldsymbol{G_c})(\mu_2 + \beta_2 \sum \boldsymbol{G_2} + \beta_{2c} \sum \boldsymbol{G_c}) \\ &= \mu_1\mu_2 + \mu_2\beta_1 \sum \boldsymbol{G_1} + \mu_1\beta_2 \sum \boldsymbol{G_2} + \beta_1\beta_2 \sum \boldsymbol{G_1} \sum \boldsymbol{G_2} + \\ &\quad \sum \boldsymbol{G_c}(\mu_1\beta_{2c} + \mu_2\beta_{1c} + \beta_1\beta_{2c} \sum \boldsymbol{G_1} + \beta_{1c}\beta_2 \sum \boldsymbol{G_2} + \beta_{1c}\beta_{2c} \sum \boldsymbol{G_c})\,. \end{aligned}$$

If we were to mean centre the risk factors and the genetic variants, then X_{12} can be expressed as:

$$\begin{aligned} X_{12} &= (X_1 - \mu_1) \times (X_2 - \mu_2) \\ &= \beta_1\beta_2 \sum \boldsymbol{G_1} \sum \boldsymbol{G_2} + \sum \boldsymbol{G_c}(\beta_1\beta_{2c} \sum \boldsymbol{G_1} + \beta_{1c}\beta_2 \sum \boldsymbol{G_2} + \beta_{1c}\beta_{2c} \sum \boldsymbol{G_c}) \,. \end{aligned} \tag{5.13}$$

By substituting Equation 5.13 and the mean centred values for X_1 and X_2 into Equation 5.5, the outcome Y can be expressed as:

$$\begin{aligned} Y = \theta_0 + \theta_1(\beta_1 \sum \boldsymbol{G_1} + \beta_{1c} \sum \boldsymbol{G_c}) + \theta_2(\beta_2 \sum \boldsymbol{G_2} + \beta_{2c} \sum \boldsymbol{G_c}) + \\ \theta_{12}(\beta_1\beta_2 \sum \boldsymbol{G_1} \sum \boldsymbol{G_2} + \sum \boldsymbol{G_c}(\beta_1\beta_{2c} \sum \boldsymbol{G_1} + \beta_{1c}\beta_2 \sum \boldsymbol{G_2} + \beta_{1c}\beta_{2c} \sum \boldsymbol{G_c})) \,. \end{aligned} \tag{5.14}$$

Consider the expected value of Y in Equation 5.14 conditional on each set of genetic variants $\boldsymbol{G_1}$, $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$:

$$\mathbb{E}(Y|\sum \boldsymbol{G_1}) = (\beta_1\theta_1 + \theta_{12}(\beta_1\beta_2 2J_2 MAF_2 + \beta_1\beta_{2c} 2J_c MAF_c)) \sum \boldsymbol{G_1} \,, \tag{5.15}$$

$$\mathbb{E}(Y|\sum \boldsymbol{G_2}) = (\beta_2\theta_2 + \theta_{12}(\beta_1\beta_2 2J_1 MAF_1 + \beta_{1c}\beta_2 2J_c MAF_c)) \sum \boldsymbol{G_2} \,, \tag{5.16}$$

$$\begin{aligned} \mathbb{E}(Y|\sum \boldsymbol{G_c}) = \beta_{1c}\theta_1 + \beta_{2c}\theta_2 + \theta_{12}(\beta_1\beta_{2c} 2J_1 MAF_1 + \beta_2\beta_{1c} 2J_2 MAF_2 + \\ 2\beta_{1c}\beta_{2c} 2J_c MAF_c) \sum \boldsymbol{G_c} \,, \end{aligned} \tag{5.17}$$

where J_1, J_2 and J_c are the number of genetic variants in $\boldsymbol{G_1}$, $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$, and MAF_1, MAF_2 and MAF_c are the MAFs of the variants. Equations 5.15 to 5.17 are approximately equivalent to the genetic associations of the outcome Y with respect to $\boldsymbol{G_1}$, $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$.

Using summarized data for the genetic associations of $\boldsymbol{G_1}$, $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$ with X_1, X_2 and Y, from Equations 5.15 to 5.17, we want to solve:

$$\begin{pmatrix} \beta_1 & 0 & A_1 \\ 0 & \beta_2 & A_2 \\ \beta_{1c} & \beta_{2c} & A_3 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_{12} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_{Y1} \\ \hat{\beta}_{Y2} \\ \hat{\beta}_{Yc} \end{pmatrix} \tag{5.18}$$

where:

$$
\begin{aligned}
A_1 &= \beta_1(\beta_2 2J_2 MAF_2 + \beta_{2c} 2J_c MAF_c)\,, \\
A_2 &= \beta_2(\beta_1 2J_1 MAF_1 + \beta_{1c} 2J_c MAF_c)\,, \\
A_3 &= \beta_1\beta_{2c} 2J_1 MAF_1 + \beta_2\beta_{1c} 2J_2 MAF_2 + 2\beta_{1c}\beta_{2c} 2J_c MAF_c\,,
\end{aligned}
$$

and $\hat{\beta}_{Y1}$, $\hat{\beta}_{Y2}$ and $\hat{\beta}_{Yc}$ represent the estimates from regressing Y against $\boldsymbol{G_1}$, $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$ respectively. We assume that there is at least one cross-over variant $\boldsymbol{G_c}$ that is associated with both risk factors, otherwise θ_1, θ_2, and θ_{12} could not be expressed as a system of three equations. By using estimates of the genetic associations β_1, β_2, β_{1c} and β_{2c}, and information on the MAFs, we want to solve $A\boldsymbol{\theta^T} = \hat{\boldsymbol{\beta}}_{\boldsymbol{Y}}$. However, we find that the matrix $\hat{A}$ is singular:

$$
\det(A) = \beta_1 \underbrace{\begin{vmatrix} \beta_2 & A_2 \\ \beta_{2c} & A_3 \end{vmatrix}}_{(1)} - 0 \begin{vmatrix} 0 & A_2 \\ \beta_{1c} & A_3 \end{vmatrix} + A_1 \underbrace{\begin{vmatrix} 0 & \beta_2 \\ \beta_{1c} & \beta_{2c} \end{vmatrix}}_{(2)}
$$

where:

$$
\begin{aligned}
(1) &= \beta_1(\beta_1\beta_2\beta_{2c} 2J_1 MAF_1 + \beta_2^2\beta_{1c} 2J_2 MAF_2 + 2\beta_2\beta_{1c}\beta_{2c} 2J_c MAF_c - \\
&\qquad \beta_1\beta_2\beta_{2c} 2J_1 MAF_1 - \beta_2\beta_{1c}\beta_{2c} 2J_c MAF_c) \\
&= \beta_1(\beta_2^2\beta_{1c} 2J_2 MAF_2 + \beta_2\beta_{1c}\beta_{2c} 2J_c MAF_c)\,, \\
(2) &= \beta_1(-\beta_2^2\beta_{1c} 2J_2 MAF_2 - \beta_2\beta_{1c}\beta_{2c} 2J_c MAF_c) \\
&\Rightarrow \det(A) = (1) - (2) = 0\,,
\end{aligned}
$$

and it is not possible to estimate $\boldsymbol{\theta^T}$ as Equations 5.15 to 5.17 are linearly dependent.

The above highlights that the multivariable IVW model cannot be extended to estimate the interaction effect between two risk factors using standard summarized level data, and this result is confirmed in the simulation study below. For illustrative purposes, the following simulations consider summarized data that would not normally be accessible from consortia, including summarized data for the product term X_{12}, and summarized data for the products of the genetic variants $\boldsymbol{G_1} \times \boldsymbol{G_2}$ with the risk factors and outcome.

Simulation study

Summarized data on the genetic associations for J genetic variants G_j $(j = 1, \ldots, J)$ with the mean centred variables X_1, X_2, X_{12}, and Y were obtained from the data generated in Section 5.4.1 when $\theta_{12} = 0.3$ and $J_1 = J_2 = J_c = 5$. The products of $\boldsymbol{G_1} \times \boldsymbol{G_2}$ were taken to create 25 new variables, and summarized data of these products with the mean centred variables X_1, X_2, X_{12}, and Y were estimated. The summarized data was used to fit the following IVW models to the 10 000 simulated datasets:

- Model 1 - estimates of the $J = 15$ genetic associations with mean centred X_1, X_2, and Y were fitted to the model in Equation 5.12.
- Model 2 - the product term $(\hat{\beta}_{X_1 j} \times \hat{\beta}_{X_2 j})$ in Equation 5.12 was replaced with the $J = 15$ genetic associations with X_{12}:

$$\hat{\beta}_{Y_j} = \theta_1 \hat{\beta}_{X_1 j} + \theta_2 \hat{\beta}_{X_2 j} + \theta_{12} \hat{\beta}_{X_{12} j} + \epsilon_j \,, \quad \text{weights} = \text{se}(\hat{\beta}_{Y_j})^{-2} \,, \tag{5.19}$$

  where $\hat{\beta}_{X_{12} j}$ are the J genetic associations with X_{12}.
- Model 3 - Equation 5.19 was re-fitted with the number of IVs increased from 15 to 40 by including summarized data on the 25 genetic products $\boldsymbol{G_1} \times \boldsymbol{G_2}$.

The following measurements were recorded for the estimate of θ_{12}: median value; relative median bias; overall bias; standard deviation; median standard error; power at the 5% significance level; and coverage.

Since this section uses the data generated in Section 5.4.1, 'summary level data' in this simulation study does not refer to the two–sample setting defined in Section 1.5.2. Instead, we have used one–sample data from Section 5.4.1 to obtain summary level data. The conclusions drawn from the simulation should not be effected by the summary level data being obtained from the one–sample rather than the two–sample setting.

Results The results from the 10 000 simulated datasets are presented in Table 5.12. For each model, the median estimate, the relative median bias of the estimate, the bias of the estimate, the standard deviation of the estimate, the median standard error, the statistical power to detect the effect at a nominal 5% significance level, and the coverage of the true effect for a 95% confidence interval are presented in Table 5.12.

When the IVW model only contained summarized data on the genetic associations with the mean centred variables X_1, X_2 and Y, and there was a product term between

$\hat{\beta}_{X_1 j}$ and $\hat{\beta}_{X_2 j}$ (Model 1), the estimates of the interaction were biased and highly unstable owing to the singular/near singular matrix in Equation 5.18. When $\hat{\beta}_{X_1 j} \times \hat{\beta}_{X_2 j}$ was replaced with the genetic associations of X_{12}, the estimates of the interaction term were consistent (Models 2 and 3). By including the genetic associations with the products $\boldsymbol{G_1} \times \boldsymbol{G_2}$ in Model 3, the power to detect the interaction increased from 4.7% (Model 2) to 23.2%.

Table 5.12 Performance of the three IVW models with respect to the interaction term using the simulated data from Section 5.4.1 when $\theta_{12} = 0.3$ and $J_1 = J_2 = J_c = 5$.

	No. of IVs	med($\hat{\theta}_{12}$)	Rel bias	Bias	SD	SE	Power	Cov
Model 1	15	0.1850	1.1369	-0.1231	0.5048	0.6930	1.1	99.1
Model 2	15	0.3048	0.6468	0.0023	0.2992	0.4050	4.7	99.6
Model 3	40	0.3088	0.5232	0.0085	0.2381	0.2474	23.2	96.0

Abbreviations: No., number; IVs, instrumental variables; Rel bias, relative median bias; SD, standard deviation; SE, standard error; Cov, coverage.

Summary

In the first half of this Section we extended multivariable Mendelian randomization with two risk factors to the factorial setting. We have demonstrated how the interaction effect of two risk factors on the outcome can be estimated using individual level data in a TSLS regression model when the genetic variants are either treated as individual IVs, or are combined into two weighted gene scores. We have shown that by mean centring the risk factors and taking the product of these variables the power to detect the main effects in TSLS regression increases. The effect of including variants that are associated with both risk factors (referred to as 'cross-over' variants) has also been considered in detail. Unlike multivariable Mendelian randomization [82], there appears to be little advantage in including genetic variants that are jointly associated with both of the risk factors to estimate the interaction effect in factorial Mendelian randomization. In our simulations, there was little evidence of weak instrument bias for the interaction term as the proportion of variance explained by the genetic variants reduced.

Through theoretical arguments and a simulation study, we have shown that it is not possible to extend the multivariable IVW model to the factorial setting using summarized data we would usually expect to obtain from a consortium. Given that the addition of an interaction term should be treated as a separate risk factor with its own IV (Section 5.2.2), this result was not surprising. As noted in Section 1.5.2, using summary level data in Mendelian randomization is becoming increasingly popular as it

is often publicly available. As such, the restriction of only being able to use individual level data may limit the scope of factorial Mendelian randomization.

In the second half of this Section we introduce a formal framework for using genetic variants as proxies for pharmacological interventions to detect statistical interactions between drug treatments. Through simulations, we consider some of the methodological challenges of this approach.

5.4.2 Genetic variants used as proxies for drug treatments

Suppose we have a biomarker X, a disease outcome Y, and a set of unmeasured confounders U of the $X - Y$ association, such that:

$$\mathbb{E}[Y|X, U] = \theta_0 + \theta_1 X + \zeta_Y U \,,$$

where ζ_Y is the effect of the unmeasured confounders on the outcome, and θ_1 is the causal effect of X on Y, with lower values of X reducing the risk of the disease outcome. Suppose there is a drug A that targets the gene region A which regulates X, and this drug reduces the risk of the disease outcome by decreasing the levels of the biomarker. If there is a genetic variant G_A in the gene region A that satisfies the IV assumptions, and the outcome Y is associated with G_A, then it can be inferred that the biomarker X causes the outcome Y. Since G_A is associated with X, and lies in the gene region that drug A is targeting, then G_A could be considered as a proxy for drug A. Hence, randomization to the genetic variant G_A acts as a proxy to being randomized to drug A. Results from a Mendelian randomization analysis that uses G_A as an IV for X could therefore provide information on the potential effect of drug A on the disease outcome.

Now suppose we have an additional drug treatment B, and this drug reduces the risk of the outcome Y by targeting a separate gene region B that also regulates X. We assume that there is a genetic variant G_B in the gene region B that satisfies the IV assumptions and is uncorrelated with G_A (not in linkage disequilibrium). Using the same logic applied to G_A, G_B could be considered as a proxy for drug B and be used in a Mendelian randomization analysis to consider the potential effect of drug B on the disease outcome. The biomarker X can be expressed in terms of G_A and G_B:

$$\mathbb{E}[X|G_A, G_B, U] = \theta_0 + \beta'_A G_A + \beta'_B G_B + \zeta'_X U \,,$$

where β'_A is the genetic association of G_A with X, β'_B is the genetic association of G_B with X, and ζ'_X is the effect of the unmeasured confounder on X.

We may suspect that there is a statistical interaction of G_A and G_B on X:

$$\mathbb{E}[X|G_A, G_B, G_{AB}, U] = \theta_0 + \beta_A G_A + \beta_B G_B + \beta_{AB} G_{AB} + \zeta_X U \,,$$

where G_{AB} represents the product term $G_A \times G_B$, and β_{AB} is the interaction effect of G_{AB} on X. If Y is regressed against G_A, G_B and G_{AB}, and there is evidence of an interaction effect, then this would suggest that the effect of reducing X via G_A on the risk of disease differs depending on whether G_B is present or not. If we assume that G_A and G_B act as proxies for drug treatments A and B, then we could infer that there is an interaction effect between drug A and drug B on the risk of disease.

To avoid issues with weak instrument bias, we may wish to identify multiple genetic variants that explain additional independent variability in the biomarker X and satisfy the IV assumptions:

$$\mathbb{E}[X|G_A, G_B, G_{AB}, U] = \theta_0 + \sum_{j=1}^{J_A} \beta_{Aj} G_{Aj} + \sum_{j=1}^{J_B} \beta_{Bj} G_{Bj} + \sum_{j_{ab}=1}^{J_A \times J_B} \beta_{ABj_{ab}} G_{ABj_{ab}} + U \,, \tag{5.20}$$

where $\boldsymbol{G_A}$ are the J_A uncorrelated genetic variants in gene region A, $\boldsymbol{G_B}$ are the uncorrelated J_B genetic variants in gene region B, and $\boldsymbol{G_{AB}}$ are the $J_A \times J_B$ product terms of $\boldsymbol{G_A} \times \boldsymbol{G_B}$. Since the $\boldsymbol{G_A}$ and $\boldsymbol{G_B}$ are selected from individual gene regions it may be difficult to identify uncorrelated genetic variants that are associated with X. As such, we would expect the number of genetic variants J_A and J_B to be modest, as seen in the applied examples considered in the literature [33–35]. Rather than treating each of the genetic variants as separate IVs, $\boldsymbol{G_A}$ and $\boldsymbol{G_B}$ could be combined into two externally weighted gene scores:

$$GS_A = \sum_{j=1}^{J_A} \beta_{Aj} G_{Aj} \qquad \text{and} \qquad GS_B = \sum_{j=1}^{J_B} \beta_{Bj} G_{Bj} \,,$$

where GS_A is the gene score for the variants in gene region A, and GS_B is the gene score for the variants in gene region B. The outcome Y could then be regressed against GS_A, GS_B and the product of the two gene scores $GS_A \times GS_B$. If there was evidence of an interaction effect between GS_A and GS_B, then under the same argument used for when $\boldsymbol{G_A}$ and $\boldsymbol{G_B}$ consisted of one variant each, we could deduce that there is an interaction effect of the drug treatments A and B on the outcome Y.

Instead of treating the gene scores as continuous variables, they could be dichotomized to create two binary variables, and these binary variables and their product could be included in a regression model with Y. Ference *et al.* [33] dichotomize the gene scores at their median value (Section 5.3.2), but rather than modelling the outcome on these dichotomized gene scores, they create a 2×2 contingency table (Table 5.2) and compare the outcome in the groups n_{10}, n_{01} and n_{11} with n_{00} in separate models.

The distribution of the weighted gene scores will affect the number of participants in each cell of the 2×2 contingency table. For the number of participants to be balanced in the contingency table the distribution of the weighted gene scores should be approximately symmetric. The MAF of the genetic variants will affect the distribution of the scores. If the genetic variants are rare, then the distribution of the gene scores will be skewed, whereas gene scores based on common genetic variants will be more symmetric. Since we are dichotomizing the weighted gene scores, the estimates of the genetic associations used as weights will also have an impact on the dichotomization. The magnitude of the genetic associations are influenced by the MAF of the variant and the amount of variance the genetic variant explains in the risk factor. Since the weights are constant values, they should not have a dramatic impact on how symmetric the distribution of the gene scores are, but they will influence the spread of the distribution.

Investigating the effect that the MAF and the proportion of variance in the risk factor explained by the genetic variants has on dichotomizing the gene scores may provide an insight into which genetic variants should be included in the gene scores. This is particularly relevant for when the genetic variants are used as proxies for drug treatments as there may be few uncorrelated genetic variants within the given gene region. For instance, we may prefer to have more common variants that explain less variation in the risk factor but the gene score is symmetric, than have rare variants that explain more variation in the risk factor and the gene score is skewed.

The simulation study below assesses the advantages of modelling the interaction term of the gene scores on the outcome, and the merit of treating the scores as continuous variables rather than dichotomizing them. As noted in Section 5.3.4, the interpretability of the estimate for the interaction term in this setting will be limited as it represents the effect of the genetic variants on the outcome, and not the effect of the treatments on the outcome. Although the simulation study will provide summary measures on the estimates of the interaction term, the main focus of the simulation to assess whether the models can detect an interaction effect. As such, the estimates of the interaction term should not be overly interpreted. Whilst the simulations do not fit the models used by Ference *et al.* [33], it does consider the impact the distribution

of the gene scores have on the numbers of participants in the 2×2 contingency table, and the effect that this has on the power to detect the interaction term.

Simulation study

Using the same notation defined above, the risk factor X was generated for $i = 1, 2, \ldots, 10\,000$ participants from the following data generating model:

$$X_i = 0.3 + \sum_{j=1}^{J_A} \beta_{Aj} G_{Aji} + \sum_{j=1}^{J_B} \beta_{Bj} G_{Bji} + \beta_{AB} \sum_{j_{ab}=1}^{J_A \times J_B} G_{ABj_{ab}i} + U_i + \epsilon_{Xi}\,.$$

We assume that the two gene regions are distinct, and the genetic variants $\boldsymbol{G_A}$ and $\boldsymbol{G_B}$ are not in linkage disequilibrium. The genotypes were generated independently from binomial distributions $\mathcal{B}(2, MAF_j)$, where MAF_j represents the MAF for the j^{th} genetic variant. MAF_j was drawn from a uniform distribution $\mathcal{U}(MAF_L, MAF_U)$, where the value of MAF_L and MAF_U differed between the two gene regions and scenarios considered. Unlike Equation 5.20, we assumed that the interaction effect β_{AB} was constant across the $J_A \times J_B$ product terms for simplicity.

The approximate proportion of variance explained in X by $\boldsymbol{G_A}$ (σ_A^2) and $\boldsymbol{G_B}$ (σ_B^2) varied between scenarios. In total, $\boldsymbol{G_A}$ and $\boldsymbol{G_B}$ explained a minimum of 6% and a maximum of 10% of the variance in X, with $\sigma_A^2 = \sigma_B^2$ or $\sigma_A^2 < \sigma_B^2$. As with the simulation study in Section 5.4.1, we acknowledge that the amount of variation explained may be too large. As done in Section 5.4.1, the genetic associations $\boldsymbol{\beta_A}$ and $\boldsymbol{\beta_B}$ were calculated by rearranging the formula for the variance of the genetic variants and ensuring the amount of variance explained by each variant was the same:

$$\beta_{Aj} = \sqrt{\frac{\sigma_A^2 / J_A}{2 \times MAF_{Aj}(1 - MAF_{Aj})}} \quad \text{and}$$

$$\beta_{Bj} = \sqrt{\frac{\sigma_B^2 / J_B}{2 \times MAF_{Bj}(1 - MAF_{Bj})}}\,.$$

As done in the previous simulation, the unmeasured confounders U were drawn from $\mathcal{N}(0, 0.25)$, and the error term ϵ_X was generated from $\mathcal{N}(0, 0.65)$. The outcome Y was generated from:

$$Y_i = \theta_0 + \theta_1 X_i + U_i + \epsilon_{Yi}\,,$$

where θ_1 represents the causal effect of X on Y, and the error term ϵ_Y was generated from a standard normal distribution $\mathcal{N}(0,1)$. The data was generated 10 000 times under the following scenarios:

- Scenario 1: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.4, 0.5)$, $\sigma_A^2 = 3\%$ and $\sigma_B^2 = 3\%$
- Scenario 2: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.4, 0.5)$, $\sigma_A^2 = 5\%$ and $\sigma_B^2 = 5\%$
- Scenario 3: $MAF_A \sim \mathcal{U}(0.1, 0.2)$, $MAF_B \sim \mathcal{U}(0.1, 0.2)$, $\sigma_A^2 = 5\%$ and $\sigma_B^2 = 5\%$
- Scenario 4: $MAF_A \sim \mathcal{U}(0.1, 0.5)$, $MAF_B \sim \mathcal{U}(0.1, 0.5)$, $\sigma_A^2 = 5\%$ and $\sigma_B^2 = 5\%$
- Scenario 5: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.1, 0.2)$, $\sigma_A^2 = 5\%$ and $\sigma_B^2 = 5\%$
- Scenario 6: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.4, 0.5)$, $\sigma_A^2 = 3\%$ and $\sigma_B^2 = 7\%$
- Scenario 7: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.1, 0.2)$, $\sigma_A^2 = 3\%$ and $\sigma_B^2 = 7\%$

with $J_A = J_B = 3$, $\theta_0 = 0.2$, $\theta_1 = 0.1$, and $\beta_{AB} = 0.1, 0.3$ and 0.5. To be consistent with the analysis performed by Ference *et al.* [33], only three genetic variants were generated for each gene region, and we chose a range of values for β_{AB} to investigate the impact this had on the results. The seven scenarios were selected to consider the impact the MAF and the amount of variance in the risk factor explained by the genetic variants had on the performance of the models. We considered genetic variants generated with a MAF$\sim \mathcal{U}(0.4, 0.5)$ to be 'common', and a MAF$\sim \mathcal{U}(0.1, 0.2)$ to be 'rare'. We acknowledge that this could be viewed as a crude classification, and the bounds for the rare variants may be considered to be too liberal.

For each scenario, optimal weighted gene scores GS_A and GS_B were generated for each gene region, where the external weights were produced from an independent set of 10 000 individuals from the same data generating model used for the main set of participants. The two gene scores were dichotomized at their median values to create two binary variables. The outcome was then regressed against: a) the two continuous gene scores and their product; and b) the dichotomized gene scores and their product. The following measurements were recorded for the estimate of the interaction effect between the gene scores on the outcome: median value; standard deviation; median standard error; and power at the 5% significance level. When the gene scores were dichotomized, the number of participants in the groups n_{00}, n_{10}, n_{01} and n_{11} were recorded.

Results

The mean numbers of participants contained in each cell of the 2×2 contingency table when the gene scores were dichotomized at the median values for each scenario are contained in Table 5.13. The results from the simulation study using 10 000 simulated datasets when the gene scores were treated as continuous and binary variables for the seven different scenarios are presented in Table 5.14. For each scenario, the median estimate, the standard deviation, the median standard error, and the statistical power to detect the effect at a nominal 5% significance level for the interaction term when the gene scores were treated as continuous and binary are presented in Table 5.14.

Table 5.13 Mean numbers (%) of participants in the groups n_{00}, n_{10}, n_{01}, and n_{11} when the gene scores were dictohomized at their median values for each scenario.

	Mean numbers (%) of participants			
Scenario	n_{00}	n_{10}	n_{01}	n_{11}
1	2,822.4 (28.2)	2,492.0 (24.9)	2,487.6 (24.9)	2,197.9 (22.0)
2	2,818.5 (28.2)	2,488.1 (24.9)	2,493.2 (24.9)	2,200.1 (22.0)
3	3,248.5 (32.5)	2,456.3 (24.6)	2,446.6 (24.5)	1,848.7 (18.5)
4	3,007.9 (30.1)	2,480.2 (24.8)	2,472.5 (24.7)	2,039.5 (20.4)
5	3,031.1 (30.3)	2,281.3 (22.8)	2,674.4 (26.7)	2,013.1 (20.1)
6	2,820.1 (28.2)	2,490.3 (24.9)	2,490.9 (24.9)	2,198.7 (22.0)
7	3,030.3 (30.3)	2,281.2 (22.8)	2,675.4 (26.8)	2,013.1 (20.1)

There was heterogeneity between the estimates of the interaction effects for the seven scenarios for the continuous and binary gene scores (Table 5.14). As we have already highlighted, the interpretability of the estimate of the interaction effect under this setting is limited, and this is confirmed by the simulation study. The introduction of the interaction term should be viewed as a means for detecting rather than estimating the interaction effect.

For all of the scenarios, the continuous gene scores had more power to detect the interaction term than the binary gene scores. The continuous and binary gene scores had the greatest power to detect the interaction effect under Scenarios 1 and 2 when both of the MAFs were generated from $\mathcal{U}(0.4, 0.5)$. Changing the values of σ_1^2 and σ_2^2 had little impact on the performance of the two models, and the power to detect the interaction effect remained the same when σ_1^2 and σ_2^2 increased from 3% (Scenario 1) to 5% (Scenario 2).

The lower and upper bounds of the uniform distributions used to generate MAF_A and MAF_B affected the power to detect the interaction term for both models. Compared

to Scenario 2 where MAF_A and MAF_B were both generated from $\mathcal{U}(0.4, 0.5)$, the power to detect the interaction term decreased as the lower and/or upper bound of the uniform distribution decreased for either both, or one of MAF_A and MAF_B. Compared to Scenario 2, the greatest reduction in power occurred when both MAF_A and MAF_B were generated from $\mathcal{U}(0.1, 0.2)$ (Scenario 3), where the power to detect the interaction term halved for the binary gene scores.

Apart from Scenarios 5 and 7, the mean values for n_{01} and n_{10} were approximately 25% (Table 5.13). The mean value for n_{00} was consistently larger than the mean value for n_{11} across all of the scenarios, and this was particularly true for Scenario 3 where $n_{00} = 32.5\%$ and $n_{11} = 18.5\%$. Note that Scenario 3 had the lowest power to detect the interaction effect for both the continuous and binary gene scores.

Recommendation

The distributions of the weighted gene scores had a significant impact on the power to detect the interaction term when the gene scores were regressed against the outcome. The results from the simulation study illustrated that the MAF has a greater influence on the performance of the models compared to the amount of variance explained by the genetic variants. This observation supports the notion that the symmetry of the distributions of the gene scores, primarily determined by the MAF, plays an important role in the performance of the models, particularly when the scores are dichotomized. If the gene scores are based on rarer variants, then the distributions of the gene scores will be skewed, whereas if the scores are based on more common variants, or the MAFs have a good range, then the distributions of the gene scores will be more symmetric. If both of the gene scores are relatively symmetric, then using the median values to dichotomize the variables will lead to more equally sized groups within the 2×2 contingency table. If there is a choice of genetic variants within each gene region, it may be advantageous to include the more common variants in the gene scores to ensure the distributions are symmetric.

Table 5.14 Estimates of the interaction between the gene scores when they were treated as continous and binary variables for the seven scenarios considered.

	Continuous gene scores				**Binary gene scores**			
	med($\hat{\theta}_{12}$)	SD	SE	Power	med($\hat{\theta}_{12}$)	SD	SE	Power
Scenario 1: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.4, 0.5)$, $\sigma^2_A = 3\%$ and $\sigma^2_B = 3\%$								
θ_{12}=0.1	0.0583	0.0420	0.0417	29.3	0.0368	0.0423	0.0421	13.5
θ_{12}=0.3	0.0330	0.0080	0.0078	98.7	0.1102	0.0429	0.0423	73.5
θ_{12}=0.5	0.0224	0.0034	0.0032	100.0	0.1846	0.0428	0.0427	98.9
Scenario 2: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.4, 0.5)$, $\sigma^2_A = 5\%$ and $\sigma^2_B = 5\%$								
θ_{12}=0.1	0.0484	0.0343	0.0343	29.1	0.0372	0.0420	0.0422	13.5
θ_{12}=0.3	0.0304	0.0074	0.0072	98.8	0.1108	0.0424	0.0423	74.3
θ_{12}=0.5	0.0212	0.0033	0.0030	100.0	0.1851	0.0439	0.0427	99.0
Scenario 3: $MAF_A \sim \mathcal{U}(0.1, 0.2)$, $MAF_B \sim \mathcal{U}(0.1, 0.2)$, $\sigma^2_A = 5\%$ and $\sigma^2_B = 5\%$								
θ_{12}=0.1	0.0824	0.1152	0.1150	10.9	0.0168	0.0435	0.0430	7.0
θ_{12}=0.3	0.1082	0.0519	0.0500	58.8	0.0526	0.0434	0.0430	23.3
θ_{12}=0.5	0.0996	0.0300	0.0278	94.6	0.0879	0.0436	0.0430	53.0
Scenario 4: $MAF_A \sim \mathcal{U}(0.1, 0.5)$, $MAF_B \sim \mathcal{U}(0.1, 0.5)$, $\sigma^2_A = 5\%$ and $\sigma^2_B = 5\%$								
θ_{12}=0.1	0.0663	0.0638	0.0603	19.9	0.0292	0.0428	0.0424	10.5
θ_{12}=0.3	0.0531	0.0218	0.0163	90.4	0.0876	0.0434	0.0425	53.6
θ_{12}=0.5	0.0410	0.0139	0.0074	100.0	0.1467	0.0463	0.0426	90.7
Scenario 5: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.1, 0.2)$, $\sigma^2_A = 5\%$ and $\sigma^2_B = 5\%$								
θ_{12}=0.1	0.0669	0.0699	0.0685	16.7	0.0246	0.0434	0.0425	9.1
θ_{12}=0.3	0.0618	0.0211	0.0204	85.5	0.0763	0.0433	0.0426	42.8
θ_{12}=0.5	0.0489	0.0109	0.0097	99.9	0.1279	0.0434	0.0428	84.1
Scenario 6: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.4, 0.5)$, $\sigma^2_A = 3\%$ and $\sigma^2_B = 7\%$								
θ_{12}=0.1	0.0498	0.0350	0.0352	29.2	0.0371	0.0422	0.0422	14.1
θ_{12}=0.3	0.0305	0.0075	0.0072	99.0	0.1106	0.0426	0.0423	74.2
θ_{12}=0.5	0.0213	0.0033	0.0030	100.0	0.1844	0.0430	0.0427	99.1
Scenario 7: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.1, 0.2)$, $\sigma^2_A = 3\%$ and $\sigma^2_B = 7\%$								
θ_{12}=0.1	0.0748	0.0756	0.0742	17.8	0.0259	0.0432	0.0426	9.7
θ_{12}=0.3	0.0649	0.0221	0.0215	85.4	0.0758	0.0430	0.0426	42.9
θ_{12}=0.5	0.0510	0.0113	0.0101	99.9	0.1271	0.0435	0.0428	83.9

Abbreviations: SD, standard deviation; SE, standard error.

5.4.3 Summary

In the second half of this Section we provided a formal framework for using weighted gene scores as proxies for drug treatments to detect interactions between pharmacological interventions. Rather than fitting multiple regression models and comparing the point estimates with no formal statistical testing, as done by Ference *et al.* [33], we propose fitting one model with an interaction term between the gene scores to detect interaction effects. However, the estimates of the interaction effects should not be overly interpreted as they represent the effect of the genetic variants on the outcome, and not the effect of the drug treatments on the outcome. In terms of the power to detect the interaction effect, the simulation study has also highlighted the benefit of treating the gene scores as continuous variables rather than binary variables.

In the next Section, we use the TSLS regression model described in Section 5.4.1 to estimate the interaction of body mass index and alcohol consumption on systolic blood pressure using data from UK Biobank. Whilst this applied example is primarily concerned with factorial Mendelian randomization analyses that use genetic variants as predictors of the risk factors, we will also apply the approach taken by Ference *et al.* [33] of dichotomizing the weighted gene scores and creating a 2×2 contingency table.

5.5 Interaction effect of body mass index and alcohol consumption on systolic blood pressure using UK Biobank data

Increased systolic blood pressure (SBP) is associated with a range of health conditions, including cardiovascular disease and diabetes [143, 144]. Identifying modifiable risk factors that are associated with SBP may help to reduce the burden of these disease outcomes in the general population. Whilst there have been numerous studies highlighting the adverse effects of increased body mass index (BMI) on SBP [145, 146], and the adverse effects of increased alcohol consumption [147], there has been little research on the combined effect of BMI and alcohol consumption on SBP. We will contribute to this gap in the literature by performing a factorial Mendelian randomization analysis using individual level data from UK Biobank to estimate the interaction effect of BMI and alcohol consumption on SBP.

All of the code for the applied example was written and performed by Jessica Rees in RStudio version 3.5.3 [86] using the packages *ivpack* [140].

5.5.1 Methods

UK Biobank is a prospective, population based cohort consisting of approximately 500,000 participants aged between 40-69 years living in the UK. Extensive baseline characteristics were collected at recruitment, including lifestyle factors, sociodemographic information, and physical attributes. For the analysis, we considered the 367,643 unrelated participants of European descent who passed data quality control measures and had genetic data.

BMI (kg/m^2) and SBP (mmHg) were measured at baseline when the participants attended the assessment centre. SBP was measured twice with at least a one minute break between measurements using a digital sphygmomanometer (Omron HEM-705IT). The average value of the two measurements were used in the analysis. Information on baseline alcohol consumption was obtained from a touchscreen questionnaire which included questions on alcohol drinking status, frequency of alcohol consumption, and beverage type. The responses to the amount of alcohol drank and beverage type were used to create a continuous variable that represented alcohol consumption in units per day. To adjust for blood pressure medication, 15 mmHg was added to SBP for individuals who reported to be on blood pressure lowering medication [148].

We used the 77 genome wide significant (p-value $< 5 \times 10^{-8}$) variants from the most recent meta-analysis by the Genetic Investigation of ANthropometric Traits (GIANT) consortium in participants of European ancestry to act as IVs for BMI [104]. For alcohol, we identified 10 genetic variants that have been shown to be associated with alcohol consumption [149], including the rs1229984 variant in the *ADH1B* gene region. The genetic variants used as IVs for BMI and alcohol consumption were cross-referenced to check for any overlap. BMI was regressed separately against each of the 10 alcohol variants, and alcohol consumption was regressed against each of the 77 BMI variants, all models were adjusted for gender, age, and the first ten principal components (PC).

Internally weighted gene scores were created for BMI GS_{BMI} based on the 77 genetic variants, and for alcohol consumption GS_{AC} based on the 10 genetic variants, and these gene scores were dichotomized at their median values to create two binary variables. A separate binary variable was generated using the rs1229984 variant only, where participants were either considered to have: a) a low alcohol consumption if they were homozygous or heterozygous for the alcohol decreasing allele; or b) a high alcohol consumption if they were homozygous for the alcohol increasing allele (as done in the paper by Carter *et al.* [136]). Using these binary variables, the following groups of participants were created:

- Low BMI, low alcohol consumption: $GS_{BMI} \leq med(GS_{BMI})$ and $GS_{AC} \leq med(GS_{AC})$ or was homozygous or heterozygous for the alcohol decreasing allele for the rs1229984 variant,
- High BMI, low alcohol consumption: $GS_{BMI} > med(GS_{BMI})$ and $GS_{AC} \leq med(GS_{AC})$ or was homozygous or heterozygous for the alcohol decreasing allele for the rs1229984 variant,
- Low BMI, high alcohol consumption: $GS_{BMI} \leq med(GS_{BMI})$ and $GS_{AC} > med(GS_{AC})$ or was homozygous for the alcohol increasing allele for the rs1229984 variant, and
- High BMI, high alcohol consumption: $GS_{BMI} \leq med(GS_{BMI})$ and $GS_{AC} > med(GS_{AC})$ or was homozygous for the alcohol increasing allele for the rs1229984 variant.

The above criteria created four groups of participants based on the dichotomized gene scores for BMI and alcohol consumption, and another four groups based on the dichotomized gene score for BMI and the rs1229984 variant. The numbers of participants, and the mean and standard deviation of BMI, alcohol consumption, and SBP were recorded for each group.

TSLS regression models of SBP were fitted to BMI, alcohol consumption, and the product of BMI and alcohol consumption. The following sets of IVs were considered:

- Model 1: the 77 variants for BMI and 10 variants for alcohol consumption, plus the 770 product terms between the two sets of variants.
- Model 2: the continuous gene scores GS_{BMI} and GS_{AC}, plus their product $GS_{BMI} \times GS_{AC}$.
- Model 3: the dichotomized gene scores of GS_{BMI} and GS_{AC}, plus their product.

The models were refitted when all of the variants for alcohol consumption were excluded apart from the rs1229984 variant. All models were adjusted for gender, age, and the first ten PCs. For each model, the estimate and standard error of the interaction term was recorded with its p-value. In total, six TSLS regression models were fitted to the dataset, and all of the models were adjusted for age, gender and the first 10 PCs. The F-statistic and the Sanderson-Windmeijer conditional F-statistic were estimated for each set of IVs with respect to BMI, alcohol consumption, and the product of BMI and alcohol consumption.

5.5.2 Results

291,781 (79.4%) of the 367,643 participants had measurements for BMI, alcohol consumption, SBP, gender and age. 57,917 (19.8%) of these 291,781 participants reported to be on blood pressure lowering medication. There were no cross-over variants between the groups of genetic variants used as IVs for BMI and alcohol consumption. rs1229984 was the only variant from the set of 10 IVs for alcohol consumption that was associated with BMI at the genome wide significance level, and 4 of the 77 variants used as IVs for BMI reached genome wide levels of significance with respect to alcohol consumption.

The numbers of participants, and mean (standard deviation) BMI, alcohol consumption and SBP by gene score classification are presented in Table 5.15. When the gene score for alcohol consumption consisted of 10 variants, the numbers of participants in the four groups were balanced. When the rs1229984 variant was used, the groups were very uneven, with each of the high alcohol consumption groups containing 47.6% of the data, suggesting that the majority of participants were homozygous for the alcohol increasing allele. There was little difference between the means and standard deviations of BMI or SBP when the IVs for alcohol consumption consisted of 1 or 10 variants. There was more discrepancy between mean alcohol consumption when only the rs1229984 variant was used, rather than the 10 genetic variants.

Table 5.15 Numbers (%) of participants and mean (standard deviation) body mass index, alcohol consumption and systolic blood pressure by gene score classification when either 10 genetic variants or the rs1229984 variant acted as IVs for alcohol consumption.

		Mean (SD)		
	No. participants (%)	BMI (kg/m^2)	Alcohol (units/day)	SBP (mmHg)
Dataset used for analysis	291,781 (100.0)	27.1 (4.51)	2.54 (2.58)	140.0 (19.8)
10 variants for alcohol				
Low BMI, low alcohol con.	73,003 (25.0)	26.6 (4.25)	2.50 (2.52)	140.6 (20.6)
High BMI, low alcohol con.	72,889 (25.0)	27.5 (4.65)	2.47 (2.50)	141.2 (20.6)
Low BMI, high alcohol con.	72,888 (25.0)	26.7 (4.30)	2.61 (2.68)	140.8 (20.7)
High BMI, high alcohol con.	73,001 (25.0)	27.6 (4.71)	2.59 (2.59)	141.3 (20.6)
rs1229984 variant for alcohol				
Low BMI, low alcohol con.	6,997 (2.4)	26.3 (4.10)	2.00 (2.04)	139.2 (20.2)
High BMI, low alcohol con.	6,863 (2.4)	27.3 (4.50)	1.95 (1.99)	139.7 (20.2)
Low BMI, high alcohol con.	138,894 (47.6)	26.7 (4.28)	2.59 (2.59)	140.8 (20.6)
High BMI, high alcohol con.	139,027 (47.6)	27.6 (4.69)	2.56 (2.56)	141.3 (20.6)

Abbreviations: No., number; SD, standard deviation; BMI, body mass index; SBP, systolic blood pressure; con., consumption.

The F-statistic and the Sanderson-Windmeijer conditional F-statistic for BMI, alcohol consumption, and the product of BMI and alcohol consumption for the different sets of IVs used in the models are contained in Table 5.16. When ten variants were

used for alcohol consumption, all of the F-statistics for Model 1 were below 10, and BMI was the only covariate that had a F-statistic greater than 10 when the rs1229984 variant was used for alcohol consumption in Model 1. The 10 genetic variants had a R^2 value of 0.28% for alcohol consumption, and the single variant had a R^2 value of 0.24%. BMI had the largest F-statistics for Models 2 and 3 for both sets of variants used for alcohol consumption. There was a significant reduction in the F-statistic for alcohol consumption and the product term between Models 2 and 3 when 10 variants for alcohol consumption were included in the gene score, whereas the reduction in the F-statistic between Models 2 and 3 when the gene scores for alcohol only consisted of one variant was minimal. Apart from Model 1 and Model 3 (10 genetic variants for alcohol consumption), the conditional F-statistic was greater than 10 for all covariates. The conditional F-statistics from Model 1 were particularly small for both sets of genetic variants for alcohol consumption.

Table 5.16 F-statistic and Sanderson-Windmeijer conditional F-statistic for body mass index, alcohol consumption and the product of body mass index and alcohol consumption by two stage least squares regression model, and the number of genetic variants used as instruments for alcohol consumption

	Model 1		**Model 2**		**Model 3**	
	F-statistic	CF-statistic	F-statistic	CF-statistic	F-statistic	CF-statistic
10 variants for alcohol						
BMI	6.8	1.3	1662.8	21.1	1054.1	7.0
Alcohol consumption	2.4	1.1	268.0	20.9	55.6	6.9
Product term[a]	2.4	1.1	298.6	21.0	73.2	6.9
rs1229984 for alcohol						
BMI	32.8	1.3	1654.9	17.2	1066.8	13.5
Alcohol consumption	7.7	1.2	245.1	17.1	241.6	13.4
Product term[a]	7.9	1.2	267.7	17.1	266.5	13.4

[a]Product of body mass index and alcohol consumption.
Abbreviations: CF-statistic, conditional F-statistic; BMI, body mass index.

Results from the TSLS regression models are contained in Table 5.17. None of the models provided evidence of an interaction effect of BMI and alcohol consumption on SBP levels. The estimates of the interaction were most precise when the variants were treated as individual IVs (Model 1), and least precise when the binary gene scores acted as IVs (Model 3). Apart from Model 3, where the standard error reduced by a third, the precision of the estimates decreased for all of the models when the number of IVs for alcohol consumption was reduced from 10 to 1. The improvement in the precision of the estimate from the dichotomized gene score model when the groups of participants were based on the different genotypes of one genetic variant

highlights the importance of considering different cut-off values for the weighted gene scores. By grouping the participants by the rs1229984 genotypes the high proportion of participants who were homozygous for the alcohol increasing allele was reflected in the model. This observation would not have been captured when the weighted gene score based on 10 genetic variants was dichotomized at the median value.

Table 5.17 Number of instrumental variables included the two stage least squares models and estimates of the interaction term with its standard error and p-value from the factorial Mendelian randomization analyses when either 10 genetic variants or the rs1229984 variant acted as IVs for alcohol consumption.

	No. of IVs	Estimate	Standard error	P-value
10 variants for alcohol				
Model 1: individual variants	857	0.0023	0.0503	0.9636
Model 2: continuous gene scores	3	0.0655	0.3402	0.8472
Model 3: binary gene scores	3	0.1011	0.6411	0.8746
rs1229984 variant for alcohol				
Model 1: individual variants	149	-0.0170	0.1136	0.8809
Model 2: continuous gene scores[a]	3	0.1917	0.3725	0.6068
Model 3: binary gene scores	3	0.1499	0.4174	0.7195

Abbreviations: No., number; IVs, instrumental variables.
[a]rs1229984 was treated as a continuous variable and could be 0, 1 and 2.

5.5.3 Summary

In this Section, we have used individual level data from UK Biobank to estimate the statistical interaction effect of BMI and alcohol consumption on SBP in a factorial Mendelian randomization study where genetic variants are used as predictors of the risk factors. None of the TSLS regression models provided evidence of an interaction effect of BMI and alcohol consumption. As seen in the simulations in Section 5.4.1, the estimate for the interaction effect was most precise when the variants were treated as individual IVs and all of the product terms were included in the analysis. Since the F-statistics for BMI and alcohol consumption were below 10 when the variants were treated as individual IVs, the estimates of the main effects from this analysis, although not presented, may have suffered from weak instrument bias. Since the simulations suggested that the interaction term does not suffer from the same weak instrument bias as the main effect terms, we should have more confidence in interpreting these estimates.

5.6 Discussion

In this Chapter, we have considered the methodological issues for factorial Mendelian randomization when the genetic variants are used as: a) predictors of risk factors to estimate interaction effects between two risk factors on an outcome; and b) are used as proxies for drug treatments to detect interaction effects between pharmacological interventions.

To estimate the interaction effect under scenario a), we expanded multivariable Mendelian randomization to the factorial setting by including an interaction term in the model. Through simulations, we have shown that consistent estimates of the interaction effect from TSLS regression can be obtained when the genetic variants are either treated as individual IVs, or as single IVs in weighted gene scores. As anticipated, the power to detect the interaction effect was noticeably higher when the variants were treated as individual instruments, and was maximized by including all of the genetic variants and products of the variants in the data generating model as IVs (referred to as the oracle model). The precision of the interaction term decreased when subsets of the product terms between the genetic variants were excluded as instruments. Reducing the amount of variance the genetic variants explained in the risk factors had little impact on the consistency of the interaction estimates when the full set of genetic variants and their products were used as instruments, although there was evidence of weak instrument bias for the main effects. We would therefore suggest that factorial Mendelian randomization be used with the primary motive of estimating the interaction effect between two risk factors.

The simulation study for scenario a) illustrated the sensitivity of the TSLS regression model to reductions in the set of product terms used as IVs compared to the data generating model. The simulation study and applied example suggested that as many of the genetic variants and cross products should be included as IVs as possible. Including large numbers of genetic variants that are not strongly associated with the risk factor in a Mendelian randomization analysis is not normally encouraged due to weak instrument bias. Although we did not observe any evidence of weak instrument bias for the interaction estimates in the simulation study, we acknowledge that many researchers would be reluctant to include hundreds of IVs in a Mendelian randomization analysis. We therefore recommend that the continuous weighted gene scores and their product be used as IVs in a sensitivity analysis. By using the gene scores as IVs, fewer parameters will be estimated, and this may help to reduce the impact of weak instrument bias. Furthermore, using the continuous gene scores in a sensitivity analysis may also provide some robustness against model misspecification.

For scenario b), we have provided a formal framework for using weighted gene scores as proxies for drug treatments to detect interactions between pharmacological interventions. If genetic variants are to be used as proxies for drug treatments to detect interactions, we suggest that the interaction effect, although not interpretable, should be estimated using continuous gene scores in a linear regression model. If the gene scores are dichotomized, then the distributions of the gene scores should be considered when determining the cut-off values used to create the binary variables.

5.6.1 Interpretation of the interaction effect

The estimate of the interaction effect of two risk factors on the outcome from a factorial Mendelian randomization analysis should be interpreted with some caution. If the genetic variants satisfy the IV assumptions for multivariable Mendelian randomization, and there truly is an interaction, then the interaction effect will have a causal interpretation. However, the interaction term may be an artefact of underlying non-linearity in the relationship between the risk factors and the outcome.

When the genetic variants are used as proxies for drug treatments, and the outcome is regressed against two gene scores and their products, the estimate of the interaction term from this model represents the effect of randomization to the genetic variants on the outcome, and not the effect of randomization to the drug treatments on the outcome. By estimating the effect of randomization to the variants on the outcome, we are essentially performing an ITT analysis.

5.6.2 Strength of the instrumental variables

We applied the recommendation of Sanderson and Windmeijer [83] of presenting the standard F-statistic and conditional F-statistic for each risk factor in a multivariable Mendelian randomization analysis to the factorial setting. Although the main effect estimates suffered from weak instrument bias when the mean F-statistic fell below 10, there was little bias in the interaction effect. As Mendelian randomization analyses can include weak IVs, it is our recommendation that under such circumstances only the interaction estimate is interpreted. Despite many of the risk factors having a mean conditional F-statistic less than 5, the estimates were not affected by weak instrument bias if the standard F-statistic was greater than 10. The role of the Sanderson-Windmeijer conditional F-statistic to measure instrument strength in factorial Mendelian randomization is not clear, and requires further consideration, particularly in relation to the conditional F-statistic for the product term X_{12}.

5.6.3 Number of cross-over variants

There are two main situations where factorial Mendelian randomization may be used: a) as a sensitivity analysis for when the primary analysis is considered to be multivariable by design; or b) as the primary analysis with the objective of estimating the interaction effect of the two risk factors. In the first case, we would expect there to be cross-over variants as risk factors in multivariable Mendelian randomization tend to be correlated, and share common genetic variants. Under the second scenario, where factorial Mendelian randomization is used as the primary analysis, it is likely that there would be no cross-over variants as the risk factors will probably be distinct. For instance, consider the example proposed by Davey Smith and Hemani [135] to estimate the interaction effect of obesity and alcohol consumption on the risk of liver disease, given that these risk factors are distinct, there will probably be few, or perhaps no genetic variants that are associated with both risk factors.

The simulations have shown that the number of cross-over variants has little impact on the estimates from the TSLS model when the full set of genetic variants and their products are included as individual IVs. The power to detect the interaction term using continuous gene scores increased, and decreased for the dichotomized gene scores, as the number of cross-over variants increased. Therefore, it may be unwise to apply the method proposed by Ference *et al.* [33] or estimate the interaction effect using the binary gene scores as IVs if there are a lot of cross-over variants. Note that when the genetic variants are used as proxies for drug treatments, there are no cross-over variants as the two groups of genetic variants are from different gene regions and are not in linkage disequilibrium, providing justification for Ference *et al's.* [33] recommendation for dichotomizing the gene scores.

5.6.4 Dichotomization of the gene scores

Throughout this Chapter, we followed the recommendation by Ference *et al.* [33] to dichotomize the gene scores at the median value. Whilst this was effective when the distribution of the gene scores were symmetric, or there were no cross-over variants, the power to detect the interaction term decreased as the distributions became more skewed, and the numbers of participants in each group of the contingency table became more unbalanced. If rare variants are included in the gene score then a different criteria for dichotomizing the gene scores should be considered, for example, if there is an obvious break in the gene scores, then perhaps this should inform the cut-off values used to create the binary variables.

5.6.5 Limitations

We have only considered continuous risk factors and continuous outcomes in the TSLS regression models, and the regression models of the outcome against the genetic variants. Only considering continuous outcome measurements could be viewed as a significant limitation. As highlighted in Section 2.3.3, the TSLS regression model will only produce approximate estimates of the odds ratio when the outcome is binary due to non-collapsibility. Since we only considered continuous outcome measurements in this Chapter, we were unable to assess the impact non-collapsibility has on the approximate measure of the odds ratio when detecting and/or estimating interaction effects. Furthermore, throughout this Chapter we have only looked at additive interactions, and since statistical interactions are scale dependent, detecting multiplicative interactions in factorial Mendelian randomization needs to be addressed. The methods proposed for factorial Mendelian randomization are also limited by only allowing for the effects of two risk factors or two drug treatments to be considered at one time.

Throughout this Chapter we have assumed that the genetic variants are uncorrelated. Whilst this may be a reasonable assumption for factorial Mendelian randomization when the genetic variants are used as predictors of the risk factor, this assumption may be restrictive for when the genetic variants are used as proxies for drug treatments as we would expect many of the variants associated with the biomarker to be correlated. Further consideration on the impact of including correlated genetic variants should therefore be considered.

We would also like to highlight some of the limitations of the simulation studies performed in this Chapter. The simulation study in Section 5.4.1 only considered misspecification of the data generating model in terms of the number of IVs included in the TSLS regression model and the impact of treating the genetic variants as one IV by generating weighted gene scores for each risk factor. Additional misspecifications, such as misspecified genetic associations, must be considered to determine the robustness of the TSLS model when estimating interaction effects. The weights generated for the gene scores in Section 5.4.1 were obtained from the same data generating model with the same number of participants used in the original simulation. Although we do not expect this assumption to hold in applied practice, and the results from the TSLS model in the simulation study using weighted gene scores may be overly precise, we made this assumption to assess the performance of Models 3 and 4 under the best possible scenario.

5.6.6 Key points from chapter

- Factorial Mendelian randomization uses genetic variants as IVs to detect interaction effects under broad two scenarios: 1) where the genetic variants act as predictors of the risk factors and are used to estimate the causal effect of two risk factors on the outcome; and 2) the genetic variants act as proxies for pharmacological interventions to identify interactions between drug treatments on the risk of disease.
- Although Factorial Mendelian randomization has already been considered in applied work, there has been little methodological developments on how this type of study design should be performed, and no attempt at estimating the interaction effect.
- We have shown that additive statistical interactions can be estimated from TSLS regression models using individual level data under a factorial Mendelian randomization study when the genetic variants are treated as individual IVs or as single IVs through weighted gene scores.
- Through simulations and an applied example, we have shown that the power to detect the interaction term in a factorial Mendelian randomization analysis is maximised when all of the genetic variants and their interactions are included as individual IVs. As a sensitivity analysis, continuous gene scores can be used as IVs to safeguard against potential bias from model misspecification or weak instruments.

Chapter 6

Effect of adiposity and body composition on asthma: A Mendelian randomization study

6.1 Introduction

This Chapter considers the main applied example of the dissertation: a Mendelian randomization study on the effect of adiposity and body composition on asthma. We partially address this research question by using body mass index (BMI) as a measure of adiposity in a univariable Mendelian randomization analysis. BMI is often considered in epidemiological studies as it only requires height and weight measurements. Although it can be measured with relative ease and little expense, BMI measurements may not accurately reflect body composition. For example, an individual may have a BMI measurement that is classed as obese, but have a perfectly healthy body composition.

To have a more comprehensive appreciation for the effect adiposity and body composition has on asthma, this Chapter also performs Mendelian randomization analyses on the effect of whole body fat mass (FM) and whole body fat free mass (FFM) on asthma. Throughout this Chapter, we will refer to 'whole body FM' as FM, and 'whole body FFM' as FFM. Since some of the genetic variants associated witth FM and FFM are the same [150], the IV3 assumption (Section 1.4) would be violated if FM and FFM were considered in separate univariable Mendelian randomization analyses. To investigate the effect of FM and FFM on asthma we therefore perform Mendelian randomization analyses under a multivariable framework, and we use the multivariable MR-Egger method developed in Chapter 4 to consider the possible violation of the

IV4 assumption for multivariable Mendelian randomization (Section 2.6.3). Since the effect of FM and FFM on asthma has not been considered in the literature using multivariable Mendelian randomization methods, the work presented in this Chapter should address this gap in the literature.

In Section 6.1.1, we discuss the importance of identifying modifiable risk factors of asthma, and highlight the benefits of performing a Mendelian randomization study on this research question. Section 6.1.2 provides detail on body composition measurements, and Section 6.1.3 reviews the literature on genome wide association studies on body composition and asthma. Section 6.1.4 provides a detailed overview of the Mendelian randomization studies considered in this Chapter. We perform a one-sample Mendelian randomization study using data from UK Biobank (Section 6.2), and a two-sample Mendelian randomization study using data from UK Biobank and the GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community) Consortium (Section 6.3). In Section 6.4, we compare the results of the Mendelian randomization studies from Sections 6.2 and 6.3.

6.1.1 Motivation

Asthma is a chronic and complex condition that can be difficult to diagnose [151]. Although the prevalence of asthma has plateaued in the UK since the 1990s [152], the NHS still spends around £1.1 billion per year on treating asthmatics [153]. Diagnosis of asthma tends to occur early on in childhood, with the number of incident cases generally decreasing with age [154]. Although the symptoms of asthma can be effectively controlled through medication, sufferers in the UK are still dying from the condition [155], and identifying modifiable risk factors of asthma should remain a priority [154].

Observational and longitudinal studies have provided evidence of a positive association between BMI and asthma in children [156–160] and adults [161–163]. Due to reverse causation and residual confounding, it is not possible to infer a causal association between BMI and asthma from these epidemiological studies. It seems likely that reverse causation may be an issue, particularly in relation to adults, as the effects of asthma may have contributed to long term physical inactivity, resulting in weight gain. The biological reasoning behind a positive association between BMI and asthma is not well understood, and various biological mechanisms have been suggested [164].

Mendelian randomization studies have considered the causal association between body composition and asthma [164, 165]. The Mendelian randomization study by Granell *et al.* [164] in children used measurements of BMI, FM and lean mass (LM)

for adiposity and body composition. These three measurements were considered in seperate univariable Mendelian randomization analyses, and there was evidence to suggest that all of these body composition measurements have a positive causal effect on asthma [164]. Since FM and LM are associated with common genetic variants [150], the IV3 assumption for univariable Mendelian randomization may have been violated for FM and LM in the paper by Granell *et al.* [164]. The study by Skaaby *et al.* [165] in adults used a reduced dataset from UK Biobank (N=162,124), and reported a positive causal association between BMI and asthma [165]. The effect of body composition on asthma has not been considered using multivariable Mendelian randomization methods, such as the multivariable IVW [28] and multivariable MR-Egger [32] methods. This Chapter addresses this gap in the literature.

6.1.2 Measurements for adiposity and body composition

BMI is derived from an individual's weight and height (BMI=weight (kg)/height^2 (m^2)), and is strongly correlated with body fat in adults [166]. Since variation in BMI measurements are usually attributed to differences in FM, an individual with a high proportion of LM or FFM may have a BMI that is classed as overweight or obese, but have a healthy body composition [167, 168]. Some epidemiological studies measure FM, FFM and LM to investigate the effect of body composition on disease outcomes. LM and FFM tend to be highly correlated as both measurements consist of muscle, bone, organs, and extracellular fluid [169]. Unlike FFM, LM also includes a small proportion of essential fats [170]. Fat mass index (FMI), fat-free mass index (FFMI), and lean mass index (LMI) can be calculated by dividing the body composition measurement (kg) by the square of height (m^2).

FM, FFM and LM can be measured through bioelectrical impedance analysis (BIA) or dual X-ray emission absorptiometry (DXA). DXA is considered to be the 'gold standard' for measuring body composition [171], but it requires specialist equipment and exposes individuals to small amounts of radiation [172]. BIA is more commonly used in clinical practice and large scale epidemiological studies as it is uses low cost, portable devices, that are simple to use [173]. BIA estimates FM and FFM from a body composition analyser that assumes each individual has a fixed water mass. BIA tends to underestimate FM in obese patients [172], and is not considered to be a reliable method for measuring FM and FFM in children [174]. Measurements for FM are generally larger under DXA than BIA, whereas FFM is usually smaller under DXA [173]. In addition to measuring FM and FFM, DXA can also be used to calculate LM. Since LM and FFM are highly correlated, many epidemiological studies only consider

one measurement. In this Chapter, we will only consider whole body FM and whole body FFM. As noted in Section 6.1, throughout this Chapter we will refer to 'whole body FM' as FM, and 'whole body FFM' as FFM.

6.1.3 Genome wide association studies on adiposity, body composition and asthma

There have been numerous GWASs on BMI in adults [175], with some of the earlier studies identifying common genetic variants in the *FTO* and *MC4R* gene regions [176, 177]. The Genetic Investigation of ANthropometric Traits (GIANT) Consortium identified 32 loci that were associated with BMI at the genome wide significance (GWS) level (p-value$< 5 \times 10^{-8}$) [178], and this figure increased to 97 when additional studies, including participants of non-European descent, were included in the analysis [104]. The 97 variants discovered by Locke *et al.* [104] explained 2.7% of the phenotypic variance of BMI, with the majority of the genetic variants lying in non-coding regions of the genome [104].

There have been few GWASs on FM or LM. Pei *et al.* [179] performed a GWAS on FM and identified 10 genetic variants that were associated with FM at the GWS level, and some of these variants had not been identified by the GIANT consortium for BMI. Zillikens *et al.* [180] performed a GWAS on LM and discovered the first set of genetic variants associated with LM at the GWS level. One of the genetic variants identified by Zillikens *et al.* [180] was not associated with any of the anthropometric traits considered in the GIANT consortium [180].

GWASs have been performed on asthma, but the variability in diagnosing the condition has made the interpretation and validation of results difficult [181]. The GABRIEL consortium was established to investigate the effect of genetic and environmental factors on the risk of asthma in the European Community [37], and one of their primary aims was to perform the largest GWAS on individuals diagnosed with asthma by a physician [37]. All of the studies provided data on individuals diagnosed with asthma before the age of 16 years, and over half of the studies provided data on adults (diagnosed with asthma 16 years or older). Five loci were associated with asthma at the GWS level using the complete dataset [37].

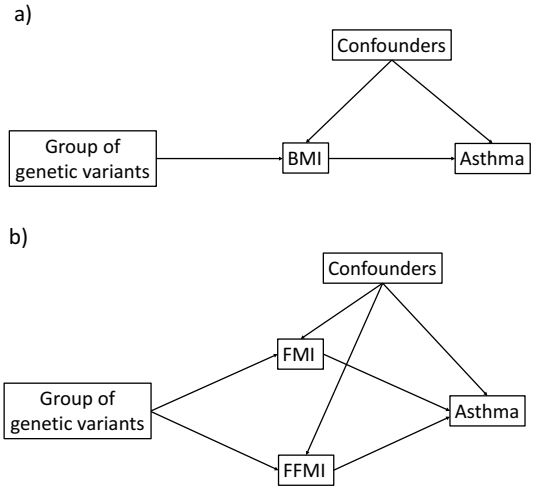
6.1.4 Study design

It could be argued that a Mendelian randomization study investigating the effect of adiposity and body composition on asthma should use data from a child cohort as

asthma tends to be diagnosed during childhood. However, cohort studies on children tend to have small sample sizes. To avoid issues with low statistical power, we therefore investigated the effect of adiposity and body composition on asthma using data on adults from UK Biobank and the GABRIEL consortium.

The directed acyclic graph (DAG) a) in Figure 6.1 illustrates the assumed relation between BMI and asthma, and we considered this DAG in univariable Mendelian randomization analyses. DAG b) in Figure 6.1 illustrates the assumed relation between FMI, FFMI and asthma, and this DAG was investigated through multivariable Mendelian randomization analyses. As highlighted by DAG b) in Figure 6.1, we assume that FM and FFM are not directly associated.

Fig. 6.1 Figure containing two directed acyclic graphs (DAG) used to inform the Mendelian randomization analyses. DAG a) illustrates the assumed relationship between body mass index (BMI) and asthma for the univariable Mendelian randomization analyses. DAG b) illustrates the assumed relationship between fat mass index (FMI), fat-free mass index (FFMI) and asthma for the multivariable Mendelian randomization analyses.



Abbreviations: BMI, body mass index; FMI, body fat mass index; and FFMI, body fat-free mass index.

For DAGs a) and b), we considered the genetic variants that reached GWS in participants of European descent for BMI in the study by Locke *et al.* [104] as potential IVs. We used the 77 genetic variants rather than the full set of 97 variants identified by Locke *et al.* [104] (as discussed in Section 6.1.3) as UK Biobank and the GABRIEL consortium predominately consist of participants of European decent. We did not consider genetic variants from GWASs on FM or FFM as IVs in any of the Mendelian randomization analyses. This decision was influenced by two factors: 1) many of the

genetic variants that effect BMI also appear to influence FM and FFM; and 2) the GWASs on BMI have significantly larger sample sizes than those on FM and FFM. Hence, the same group of genetic variants were used as IVs in DAGs a) and b) in Figure 6.1.

To investigate the effect of adiposity and body composition on asthma using the 77 genetic variants identified by Locke *et al.* [104] as potential IVs, we performed two Mendelian randomization studies: a one–sample Mendelian randomization study using data from UK Biobank (Section 6.2); and a two–sample Mendelian randomization study using data from UK Biobank and the GABRIEL consortium (Section 6.3). Figure 6.2 provides an overview of the characteristics of these studies. Both of these Mendelian randomization studies performed univariable and multivariable analyses to investigate DAGs a) and b) in Figure 6.1.

Fig. 6.2 Figure highlighting the characteristics of the two Mendelian randomization studies performed in Sections 6.2 and 6.3.

Study 1 (Section 6.3)	**Study 2 (Section 6.4)**
Study design: one–sample Mendelian randomization	**Study design**: two–sample Mendelian randomization
Risk factors: 1) BMI (univariable analyses); and 2) FMI and FFMI (multivariable analyses)	**Risk factors**: 1) BMI (univariable analyses); and 2) FMI and FFMI (multivariable analyses)
Data used: Summary level[a]	**Data used**: Summary level[a]
Source of data: Summary level data for BMI, FMI, FFMI and asthma were estimated by Jessica Rees using data from UK Biobank	**Source of data**: Summary level data for BMI, FMI and FFMI were estimated by Jessica Rees using data from UK Biobank. Summary level data for asthma was taken from the GABRIEL consortium.

[a]Note that summary level data was used in the analysis for the one–sample (Study 1) and two–sample (Study 2) Mendelian randomization studies.
Abbreviations: BMI, body mass index; FMI, body fat mass index; and FFMI, body fat-free mass index.

Summary level data is normally used in a two–sample Mendelian randomization study when the genetic associations with the risk factor and the genetic associations with the outcome are obtained from two separate samples (Section 1.5.2). Although a one–sample Mendelian randomization study typically uses individual level data in

TSLS regression, it is possible for summary level data to be used in the analysis (Figure 1.2). As stated in Section 1.5.2, throughout this dissertation we have assumed that 'summary level data' refers to the two–sample setting. However, the one–sample Mendelian randomization study performed in Section 6.2 uses summary level data in the univariable (IVW, median and MR-Egger) and multivariable (IVW and MR-Egger) analyses. The summary level data for BMI, FMI, FFMI and asthma were estimated by Jessica Rees using data on the same set of participants from UK Biobank (more detail provided in Section 6.2).

Weak instruments in a one–sample Mendelian randomization study can bias the estimate towards the confounded observational estimate when individual level data is used in the TSLS regression model [182]. However, the one–sample Mendelian randomization study considered in this Chapter applies summary level data to methods specifically designed for the two–sample setting (IVW, median and MR-Egger). Since the estimates from the IVW and TSLS methods are asymptotically equivalent, the degree of weak instrument bias for the IVW method using one–sample summary level data should be approximately equivalent to the weak instrument bias for the TSLS method in finite samples [182]. Hartwig *et al.* [183] highlighted that weak instrument bias for univariable MR-Egger under the one–sample setting will bias the estimate towards the confounded observational estimate.

Since the one–sample Mendelian randomization study in Section 6.2 may suffer from an increased Type I error rate, we also performed a two–sample Mendelian randomization study as weak instruments under the two–sample setting will bias the estimate towards the null. For the two–sample Mendelian randomization study in Section 6.3, summary level data was used under the more conventional setting of obtaining the genetic associations with the risk factors (BMI, FMI and FFMI) and the genetic associations with the outcome (asthma) from two separate samples. The summary level data for BMI, FMI and FFMI estimated in Section 6.2 using data from UK Biobank was used in the two–sample setting, and summary level data for asthma was extracted from the GABRIEL consortium.

6.1.5 Summary

In this Section, we have motivated the main applied example of the dissertation: the investigation into the effect of adiposity and body composition on asthma using Mendelian randomization. We have highlighted the benefits of performing both univariable and multivariable Mendelian randomization analyses. We have also provided

an in depth overview of the two Mendelian randomization studies considered in Sections 6.2 and 6.3.

6.2 One–sample Mendelian randomization analysis on the effects of adiposity and body composition on asthma

This Section contains the one–sample Mendelian randomization study on the effects of adiposity and body composition on asthma. BMI was used as a measure for adiposity, and its effect on asthma was considered in univariable Mendelian randomization analyses. The effect of body composition on asthma was considered through multivariable Mendelian randomization with FMI and FFMI used as measures for body composition. To address these research questions, we considered two criteria for asthma status: 1) ever diagnosis; and 2) current diagnosis (details provided in Section 6.2.1).

Methods

As highlighted in Section 6.1.4, summary level data was used in all of the analyses for this one–sample Mendelian randomization study, with individual level data from UK Biobank used to generate the summary level data. UK Biobank is a prospective cohort study of approximately 500,000 participants aged between 40-69 years who were recruited across 22 centres in the UK between 2006 and 2010. The study was established to promote research into the aetiology of common diseases by environmental, lifestyle, and genetic components [184]. Interview led questionnaires, physical measurements and touch screen questionnaires were used to collect extensive baseline characteristics at recruitment, including lifestyle factors, socio-demographic information, medical history, biological samples, and physical attributes [185].

In Section 6.2.1 we provide an overview of the exposure and outcome measurements recorded by UK Biobank that were used to generate the summary level data for the Mendelian randomization analyses. Section 6.2.2 outlines the quality control steps applied to the genetic and phenotypic data. The criteria applied to the 77 genetic variants that reached GWS in the study by Locke *et al.* [104] to select the IVs for the Mendelian randomization analyses is discussed in Section 6.2.3. Finally, we outline the models fitted to obtain the summary level data on the selected IVs (Section 6.2.4), and the Mendelian randomization analyses performed on the summary level data (Section 6.2.5).

6.2.1 Exposure and outcome measurements

Participants were asked a range of questions about their medical history in the touchscreen questionnaire at recruitment, including the question (data field 6152): 'Has a doctor ever told you that you have had any of the following conditions? (You can select more than one answer) a) blood clot in the leg; b) blood clot in the lung; c) emphysema/chronic bronchitis; d) asthma; e) hayfever, allergic rhinitis or eczema; f) none of the above' [186]. Participants were also asked whether they had experienced any wheezing or whistling in the chest during the last year. The participants underwent a verbal interview by a trained nurse to confirm whether they had provided the correct answers to the medical history questions in the touchscreen questionnaire [187]. If the participant had selected the wrong response, then their answers were amended by the interviewer, and recorded in data field 20002.

Since question 6152 included the phrase 'has a a doctor *ever* told you', and many children outgrow asthma, we were concerned that a large proportion of the UK Biobank participants who answered d) to question 6152 were not asthmatic at baseline. As such, we considered two definitions for asthma: 1) ever diagnosis of asthma; and 2) current asthma. For definition 1), a participant was considered as asthmatic if they included d) in their response to question 6152 following their interview with the trained nurse (recorded in 20002). For definition 2), a participant was considered as a current asthmatic if they had been classed as having an ever diagnosis of asthma, and had reported wheezing or whistling in the chest during the last year. We anticipated that participants classed as having 'current asthma' were more likely to be asthmatic at baseline.

If the participant had included d) in their response to question 6152, they were asked what age they were when the asthma was first diagnosed. The mean (standard deviation) age of diagnosis was calculated, and the number of participants who were diagnosed with asthma by age 9 recorded. The dataset was split at the median year of birth, and the mean (standard deviation) age of diagnosis for asthma was calculated in the two subgroups. During the physical assessment, forced volume vital capacity (FVC) and forced expiratory volume in 1 second (FEV1) were measured by the participant blowing into a spirometer two or three times [188], and the average value of the these measurements were recorded.

Weight (kg) was measured at recruitment during the physical assessment using the Tanita BC418MA body composition analyser, and standing height (m) was measured using the Seca 202 device [188]. Whole body FM (kg) and whole body FFM (kg) were estimated from bioimpedance analysis (BIA) using the Tanita BC418MA body

composition analyser [188]. As stated in Section 6.1, we will refer to 'whole body FM' as 'FM', and 'whole body FFM' as 'FFM'. FM and FFM were also estimated in a subset of participants (approximately 5,000 people) in a pilot study using Dual-energy X-ray (DXA) data from a GE-Lunar iDXA [189]. Weight, FM and FFM were converted into indices by dividing each measurement by the square of the participant's height. For those participants who had measurements for both BIA and DXA, the average difference between the measurements for fat mass index (FMI=fat mass (kg)/height^2 (m^2)) and fat-free mass index (FFMI=fat-free mass (kg)/height^2 (m^2)) was recorded.

Summary statistics of demographic characteristics, including body composition measurements and asthma status, were recorded for the participants whose measurements were used to create the summary level data for the Mendelian randomization analyses. The distributions of height, weight, BMI, FMI (both BIA and DXA measurements) and FFMI (both BIA and DXA measurements) were also assessed through histograms.

6.2.2 Data quality control

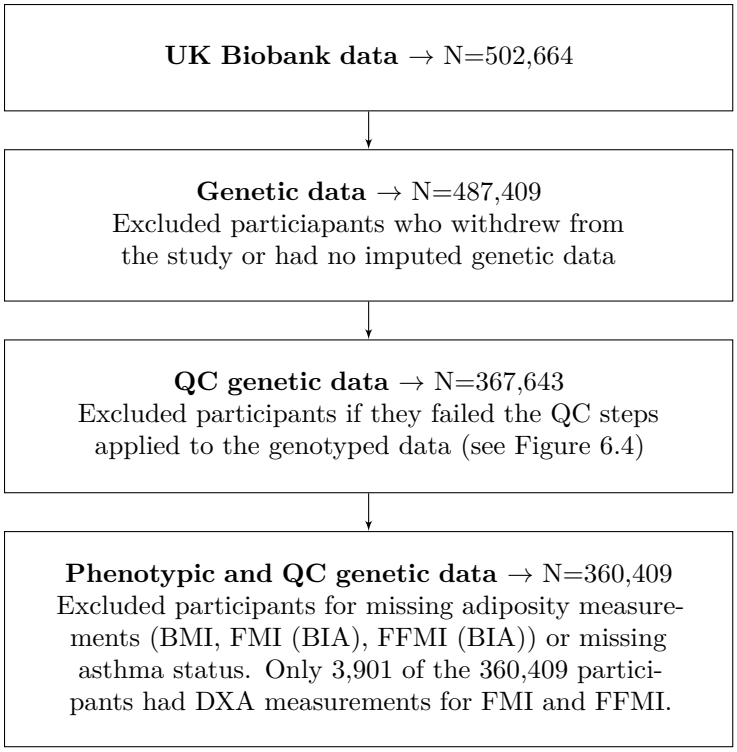
This Mendelian randomization study uses genetic data released by UK Biobank in July 2017. Genome-wide genotyping was carried out by UK Biobank in 106 batches of samples using the Affymetrix UK BiLEVE Axiom array or the Affymetrix UK Biobank Axiom array for 805,426 genetic variants. Imputation was performed centrally by the Wellcome Trust Centre for Human Genetics using the Haplotype Reference Cohort.

After excluding participants who had withdrawn from the study or had no imputed genetic data, genotyped data was available on 487,409 participants of the initial 502,664 participants (Figure 6.3). The quality control (QC) steps applied to the genotyped data on the autosomal chromosomes by the Cardiovascular Epidemiology Unit at the University of Cambridge are outlined in Figure 6.4. 782,205 genotyped genetic variants on 367,643 participants passed the QC steps. Participants with missing data for BMI, FMI (BIA measurement), FFMI (BIA measurement), ever diagnosis of asthma or current asthma were excluded from the 367,643 participants, leaving 360,409 participants with phenotypic (BMI, FMI (BIA), FFMI (BIA), and asthma status) and QC genetic data (Figure 6.3). Only 3,901 of the 360,409 participants had DXA measurements for FMI and FFMI.

As outlined in Section 6.1.4, the 77 genetic variants that reached GWS for BMI in the study by Locke *et al.* [104] were considered as potential IVs. All of the 77 genetic variants passed the QC steps outlined in Figure 6.4. Regardless of whether genotyped data was available, imputed genetic data was used to generate the summary level data for the Mendelian randomization analyses (see Section 6.2.4). Jessica Rees used the

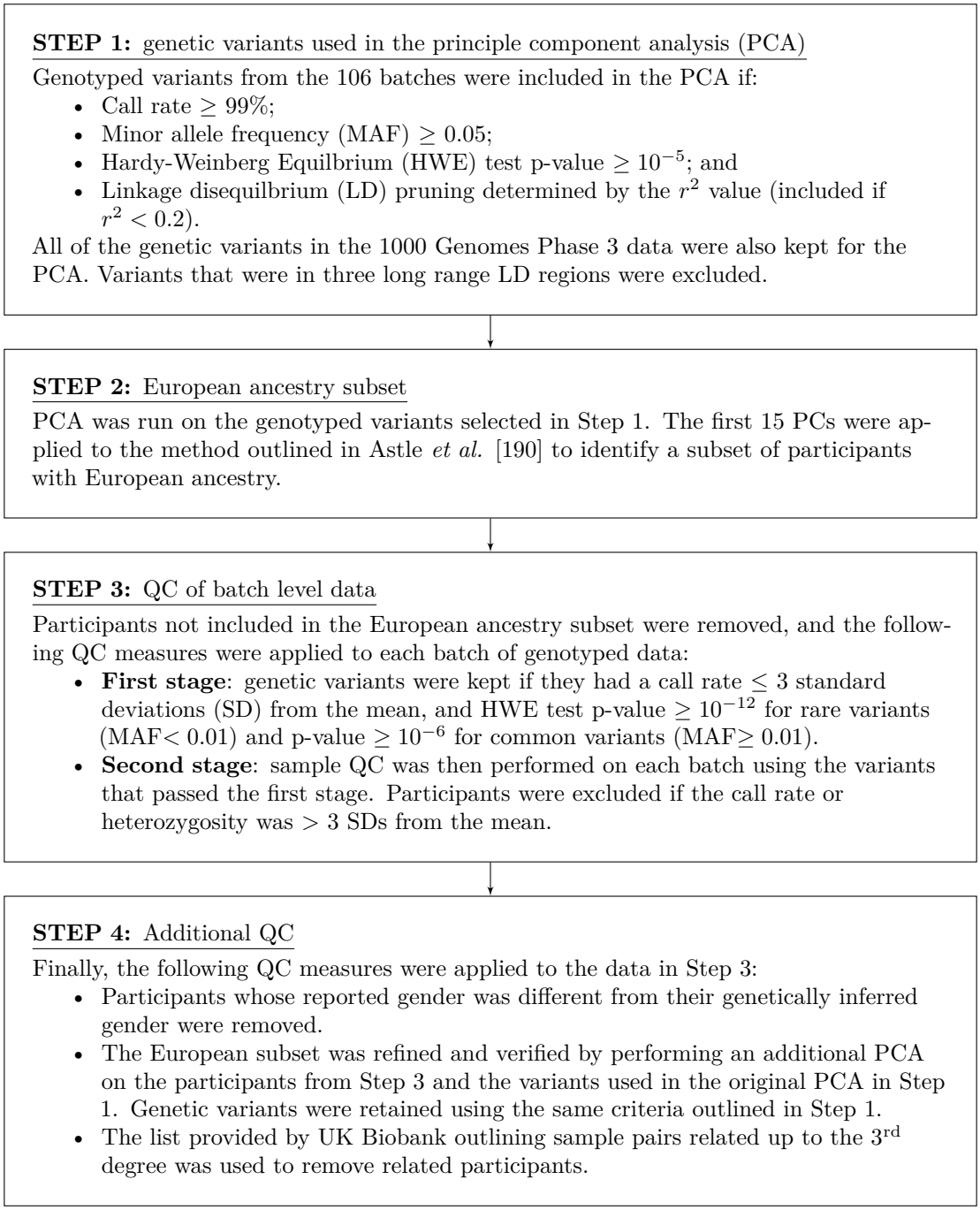
QCTOOL v2 command–line programme to extract imputed genetic data on the 77 variants for the 360,409 participants. All of the genetic variants had an imputation quality score > 0.96 (Table I.1).

Fig. 6.3 Flowchart highlighting the number of participants with phenotypic and quality controlled (QC) genetic data from UK Biobank.



Abbreviations: BMI, body mass index; FMI, body fat mass index; and FFMI, body fat-free mass index; BIA, bioelectrical impedance analysis; DXA, dual X-ray emission absorptiometry.

Fig. 6.4 Flowchart highlighting the quality control steps applied to the genotyped genetic data from UK Biobank by the Cardiovascular Epidemiology Unit at the University of Cambridge.



6.2.3 Selection of genetic variants

To identify possible pleiotropic variants, the rs numbers of the 77 genetic variants that reached GWS for BMI in the study by Locke *et al.* [104] were searched in Phenoscanner [26] with the option of searching for genetic variants that could act as proxies for the 77 genetic variants. Phenoscanner identified proxies for the 77 genetic variants by using a pairwise r^2 (based on the 1,000 Genome Project) threshold of 0.6. Hence, a genetic variant was classed as a proxy variant if its r^2 value with one of the specified 77 genetic variants was greater than 0.6. Note that the r^2 value takes into account the linkage disequilibrium between the genetic variants and the minor allele frequency of the genetic variants.

Data was extracted from the Phenoscanner search on all of the genetic variants that reached GWS with any trait. A function written by Jessica Rees in RStudio version 3.5.3 [86] was applied to the extracted data to create a list of traits (excluding BMI, weight, height, FM and FFM) for each of the 77 genetic variants that were associated at the GWS level with either the variant itself, or a proxy of the variant. These lists were recorded for each of the 77 genetic variants by whether the trait was adiposity related or not (Table 6.1). 29 of the 77 variants were associated with at least one non-adiposity related trait at GWS, and 13 variants were associated with age at menarche at GWS.

Since socioeconomic status and smoking may act as confounders of the BMI-asthma association, the p-values of the genetic associations of the 77 variants with years of educational attainment from the Social Science Genetic Association Consortium (SSGAC) [191], and with smoking status from the Tobacco, Alcohol and Genetics (TAG) consortium [192] were extracted from Phenoscanner (Table 6.1). Two variants were associated with years of educational attainment at GWS, and 32 were nominally associated (p-value > 0.5) with years of educational attainment. None of the variants were associated with smoking at GWS, whereas 13 were associated at the nominal level.

Physical activity may also act as a confounder of the BMI-asthma association, and since there was no data for this trait in Phenoscanner, the genetic associations of the 77 variants with duration of walks (minutes per day) and duration of moderate activity (minutes per day) were estimated in the UK Biobank dataset using the imputed genetic data. Jessica Rees used the command–line programme *SNPTEST v2* with the options *-method expected*, *-frequentist 1*, and *-use_raw_phenotypes* to estimate these genetic associations adjusted for the first 10 ancestry informative principle components (PC) (obtained by the Cardiovascular Epidemiology Unit in Step 4 of Figure 6.4) and gender.

By specifying '*-method expected*' as an option, *SNPTEST v2* uses the three genotype probabilities (with respect to the minor allele) to calculate the expected genotype for each variant. Note that this method does not take into account the uncertainty in the imputed genotypes. The expected genotype is applied to an additive linear regression model (*-frequentist 1*) with the option of mean centering and scaling the phenotype turned off. None of the genetic associations for duration of walking or duration of moderate exercise estimated from the UK Biobank dataset using *SNPTEST v2* reached GWS.

Using the information in Table 6.1, two sets of genetic variants were created using the following criteria: 1) under the liberal approach, the variants that were associated with smoking status at the nominal significance level (p-value < 0.05), or associated with years of educational attainment or physical activity at the GWS level (p-value $< 5 \times 10^{-8}$), or associated with five or more non-adiposity related independent traits from Phenoscanner at the GWS level were excluded; and 2) in addition to the variants excluded under the liberal criteria, the conservative approach excluded variants that were associated with years of educational attainment at the nominal level.

Note that the conservative set of genetic variants was a subset of the liberal set of genetic variants, and the two sets only differed by whether the genetic variant was nominally associated with years of educational attainment or not. A nominal level of significance for years of educational attainment was only considered for the conservative set of genetic variants as educational attainment was used as an approximate measure for socioeconomic status. However, a GWS level for educational attainment was considered for the liberal set of variants.

In total, 60 variants passed the liberal criteria, and 39 passed the conservative criteria (see Table I.1). The Mendelian randomization analyses outlined in Section 6.2.5 were applied to both sets of genetic variants. Note that the rs1558902 variant in the *FTO* gene region was associated with the following six traits at the GWS level: type 2 diabetes, age at menarche, fasting insulin, dietary macronutrient intake, HDL-C and metabolic syndrome. Under the liberal criteria this genetic variant should have been excluded from the analysis. However, many of these six traits are related to BMI and are not strictly independent (e.g. type 2 diabetes, fasting insulin). Since the six traits have no obvious biological relationship with asthma, it seems unlikely that the IV3 assumption for univariable Mendelian randomization (Section 1.4) and the IV4 assumption for multivariable Mendelian randomization (Section 2.6.3) would be violated. As such, the rs1558902 variant was not excluded from the liberal or conservative sets of variants used in the Mendelian randomization analysis.

Table 6.1 Genetic variants from the 77 variants selected from the GIANT consortium [104] that were associated with a phenotype at the genome wide significance level classified as non-adiposity or adiposity related, and/or associated with years of education and smoking at the nominal significance level (p-values provided).

rs no.	**Chr no.**	**Gene**	**GWS**[a]		**P-value**	
			Non-adiposity	Adiposity	Years of education[b]	Ever smoker[c]
rs11165643	1	*PTBP2*		HC		
rs11583200	1	*ELAVL4*			0.04237	0.02848
rs12401738	1	*FUBP1*		HC	0.00002	
rs12566985	1	*FPGT-TNNI3K*	Age at menarche	HC		
rs17024393	1	*GT2*		HC		
rs2820292	1	*V1*		WC	0.00180	
rs3101336	1	*NEGR1*	Age at menarche	HC	2.09×10^{-6}	0.01734
rs543874	1	*SEC16B*	Age at menarche	HC, BFP	0.04843	
rs657452	1	*AGBL4*		WC		
rs1016287	2	*LINC01122*		HC	0.02463	
rs10182181	2	*ADCY3*		HC		0.03967
rs11126666	2	*KCNK3*			0.02694	
rs13021737	2	*TMEM18*	Age at menarche, stretch marks	HC, BFP		0.02040
rs1528435	2	*UBE2E3*			0.01873	
rs2121279	2	*LRP1B*				0.02124
rs13078960	3	*CADM2*		HC		

rs1516725	3	*ETV5*	Age at menarche, log(eGFR creatinine in non diabetics)	HC	0.00004	
rs16851483	3	*RASA2*			0.00466	
rs2365389	3	*FHIT*			0.04086	
rs3849570	3	*GBE1*		WC		
rs6804842	3	*RARB*				
rs10938397	4	*GNPDA2*	Age at menarche	HC	0.00526	
rs13107325	4	*SLC39A8*	HDL-C, Chron's disease, schizophrenia, blood pressure, DBP, SBP, arterial pressure		0.00683	
rs2112347	5	*POC5*	LDL-C, total cholesterol	HC	0.00068	
rs13191362	6	*PARK2*	Methylation levels		0.02458	
rs2033529	6	*TDRG1*		WC		
rs205262	6	*C6orf106*	HDL-C	HC	0.01071	
rs2207139	6	*TFAP2B*		HC		
rs9400239	6	*FOXO3*	Schizophrenia	WC	0.02136	0.04902
rs1167827	7	*HIP1*			0.02364	
rs2245368	7	*PMS2L11*			0.03428	
rs17405819	8	*HNF4G*			0.00189	
rs10968576	9	*LINGO2*		WC		
rs1928295	9	*TLR4*		WC	0.01243	
rs4740619	9	*C9orf93*			0.04199	
rs6477694	9	*EPB41L4B*				
rs11191560	10	*NT5C2*	CAD, schizophrenia, SBP			
rs7903146	10	*TCF7L2*	Type 2 diabetes, fasting glucose and insulin	HC		

rs11030104	11	*BDNF*	Age at menarche	HC	0.04650	0.00020
rs12286929	11	*CADM1*			0.03439	
rs3817334	11	*MTCH2*	log(Proinsulin), HDL-C	WC		
rs4256980	11	*TRIM66*	Age at menarche	WC		
rs11057405	12	*CLIP1*	HDL-C, adiponectin levels	HC		
rs7138803	12	*BCDIN3D*	Age at menarche	HC		
rs12429545	13	*OLFM4*		HC		
rs9581854	13	*MTIF3*		Weight		0.02130
rs10132280	14	*STXBP6*		HC		
rs11847697	14	*PRKD1*				
rs12885454	14	*PRKD1*		WC		
rs7141420	14	*NRXN3*		HC	0.04606	
rs16951275	15	*MAP2K5*	Age at menarche	WC		0.03644
rs12446632	16	*GPRC5B*	Age at menarche	HC		
rs1558902	16	*FTO*	Type 2 diabetes, age at menarche, fasting insulin, dietary macro-nutrient intake, HDL-C, metabolic syndrome	HC, BFP, WC		
rs2650492	16	*SBK1*	Crohn's disease, inflammatory bowel disease, methylation levels	HC	0.02098	0.01912
rs3888190	16	*ATP2A1*	Crohn's disease, inflammatory bowel disease, years of educational attainment, methylation levels, exon level expression of *TUFM*	HC, BFP	6.96×10^{-8}	
rs758747	16	*NLRC3*			0.00739	
rs9925964	16	*KAT8*	Triglycerides, warfarin dose, Parkinson's	HC		

rs1000940	17	*RABEP1*	Age at menopause, differential exon level expression of *RPAIN*, splicing of *RABEP1*			
rs1808579	18	*C18orf8*	Years of educational attainment	HC	2.15×10^{-8}	
rs6567160	18	*MC4R*	Type 2 diabetes, HDL-C, CAD	HC, BFP		
rs17724992	19	*GDF15*		WC	0.00031	0.02425
rs2075650	19	*TOMM40*	HDL-C, LDL-C, total cholesterol, triglycerides, CAD, macular degeneration, Alzheimer's disease, cognitive decline, longevity, C-reactive protein, *APOB* apolipoprotein, cerebrospil fluid, neurofibrillary tangles, neuritic plaque, verbal memory, cortical amyloid beta load	WC, BFP	0.01957	
rs2287019	19	*QPCTL*	Fasting blood glucose, insulinogenic index, LDL-C, total cholesterol	HC		0.04889
rs29941	19	*KCTD15*			0.04850	0.03473
rs3810291	19	*ZC3H4*		HC		

Abbreviations: no., number; Chr, chromosome; GWS, genome wide significance; BFP, Body fat percentage; CAD, Coronary artery disease; DBP, Diastolic blood pressure; HC, Hip circumference; HDL-C, High-density lipoprotein cholesterol; LDL-C, Low-density lipoprotein cholesterol; SBP, Systolic blood pressure; WC, Waist circumference. [a]Genome wide significance, p-value$< 5 \times 10^{-8}$. [b]Taken the Social Science Genetic Association Consortium from Phenoscanner [191]. [c]Taken from the Tobacco, Alcohol and Genetics consortium from Phenoscanner [192].

6.2.4 Summary level data

Summary level data for the 60 genetic variants selected under the liberal criteria (Section 6.2.3) was obtained by Jessica Rees using *SNPTEST v2*. For each genetic variant, genetic associations were estimated for BMI, FMI (both BIA and DXA measurements), FFMI (both BIA and DXA measurements), and both definitions of asthma status (ever diagnosis and current asthma). The genetic associations for BMI, FMI (BIA), FFMI (BIA) and asthma (ever diagnosis and current asthma) were estimated using data on 360,409 participants. The genetic association for DXA measurements of FMI and FFMI were estimated using data on 3,901 of the 360,409 participants (Figure 6.4).

As done in Section 6.2.3 for the physical activity measurements, *SNPTEST v2* was used to estimate the genetic associations with BMI, FMI (BIA and DXA measurements) and FFMI (BIA and DXA measurements) with the options *-method expected*, *-frequentist 1*, and *-use_raw_phenotypes*. Hence, *SNPTEST v2* fitted linear regression models where the effect of the genetic variant on the body composition measurement was assumed to be additive for each additional copy of the minor allele. The genetic associations for BMI, FMI (BIA and DXA measurements) and FFMI (BIA and DXA measurements) were adjusted for population stratification by including the first 10 ancestry informative PCs in the linear regression models (obtained by the Cardiovascular Epidemiology Unit in Step 4 of Figure 6.4). The genetic associations were also adjusted for gender to improve the precision of the estimates. Despite accounting for height in the formulae for BMI, FMI and FFMI (Section 6.2.1), the measurements may not be fully independent of height. In addition to the first 10 PCs and gender, we therefore decided to adjust the BMI, FMI and FFMI genetic associations by height.

For linear regression models, we assume that the residuals are independent, normally distributed, and homoscedastic. Since the distribution of the measurements for BMI, FMI (BIA and DXA measurements) and FFMI (BIA and DXA measurements) were bimodal with respect to gender, and tended to be positively skewed (Figures I.1-I.3), the normality and homoscedasticity assumptions may not have been satisfied. Although this may have been rectified by transforming the BMI, FMI and FFMI data, the estimates from the Mendelian randomization analyses would not have been as easily interpreted. As such, the *-use_raw_phenotypes* option was used to ensure BMI, FMI and FFMI were not mean centred or scaled. The genetic associations obtained from *SNPTEST v2* therefore represent the average change in BMI, FMI or FFMI per additional copy of the minor allele.

The power to detect a genetic association with a continuous trait was estimated by Jessica Rees using the function *GeneticPower.Quantitative.Numeric()* in the R package *GeneticsDesign* [193]. Assuming an additive effect of the genetic variant on the trait, and applying a GWS level, there would be 50.7% power to detect a genetic effect if the variant explained 0.01% of the variance in the trait. This power calculation is based on a sample size of 360,409 and the assumption that 13 additional parameters (10 PCs, gender, age and height) are included in the linear regression model.

Since 60 genetic associations will be estimated on each of the BMI, FMI and FFMI measurements, using a GWS level for power calculations may be conservative. As such, it may be more appropriate to use a Bonferonni adjusted significance level of 0.05/60. Under this significance level, and using the same assumptions as above, there would be 99.0% power to detect a genetic effect. This power was reduced to 0.002% when the sample size was 3,901 rather than 360,409.

As a measure of instrument strength, the R^2, adjusted R^2 and F-statistic were recorded when BMI, FMI (BIA and DXA measurements) and FFMI (BIA and DXA measurements) were regressed against each of the 60 genetic variants in turn, and when they were regressed against all of the 60 liberal and 39 conservative variants. To consider the strength of the genetic variants for the multivariable Mendelian randomization analyses, the Sanderson-Windmeijer conditional F-statistic [83] were estimated for all of the BIA and DXA measurements for the liberal and conservative sets of genetic variants.

For asthma status (ever diagnosis and current asthma), all of the genetic associations were adjusted for the first 10 ancestry informative PCs and gender with the *SNPTEST v2* options *-method expected* and *-frequentist 1.* Hence, *SNPTEST v2* fitted logistic regression models where the effect of the genetic variant on asthma status was assumed to be additive on the log scale. Note that the logistic regression models included the first 10 ancestry informative PCs to adjust for population stratification. The genetic associations were also adjusted for gender to improve the precision of the estimates.

Plots of the genetic associations with 95% confidence intervals with BMI and asthma were assessed for pleiotropic variants. These plots were created by Jessica Rees in RStudio version 3.5.3 [86] using the package *ggplot2* [141]. To consider collinearity in the multivariable Mendelian randomization models (as discussed in Section 2.6.3), the correlation between the genetic associations of FMI and FFMI were recorded for the BIA and DXA measurements. Plots of the genetic associations with 95% confidence intervals with FMI and FFMI were also considered for the BIA and DXA measurements.

6.2.5 Mendelian randomization analyses

The summary level data obtained from the UK Biobank dataset (Section 6.2.4) was used to perform univariable and multivariable Mendelian randomization analyses (Figure 6.5). These analyses were performed on the genetic associations for an ever diagnosis of asthma and current asthma (Section 6.2.1). As outlined in Figure 6.5, the effect of BMI on asthma (ever diagnosis and current) was considered through IVW models using the following three groups of IVs: 1) 60 liberal genetic variants selected in Section 6.2.3; 2) 39 conservative genetic variants selected in Section 6.2.3; and 3) two genetic variants from the *FTO* and *MC4R* gene regions (i.e. one genetic variant from each region). Groups 1) and 2) were considered in the following sensitivity analyses: simple median [52]; weighted median [52]; and MR-Egger [29].

Fig. 6.5 Figure highlighting the univariable and multivariable Mendelian randomization analyses performed to investigate the effect of adiposity and body composition on asthma.

Univariable analyses

Risk factor: BMI

Methods: univariable IVW, median estimator and univariable MR-Egger using summary level data (Section 6.2.4)

IVs used in the methods: set of liberal genetic variants (60 IVs), set of conservative genetic variants (39 IVs), and 2 genetic variants from gene regions *FTO* and *MC4R* (applied to IVW only)

No. of analyses: 7 for ever diagnosis of asthma and 7 for current asthma

Multivariable analyses

Risk factors: FMI and FFMI (BIA and DXA measurements considered separately)

Methods: multivariable IVW and multivariable MR-Egger using summary level data (Section 6.2.4)

IVs used in the methods: set of liberal genetic variants (60 IVs) and set of conservative genetic variants (39 IVs)

No. of analyses for BIA and DXA measurements: 8 for ever diagnosis of asthma and 8 for current asthma

Abbreviations: BMI, body mass index; FMI, body fat mass index; and FFMI, body fat-free mass index; BIA, bioelectrical impedance analysis; DXA, dual X-ray emission absorptiometry; IVW, inverse variance weighted; IV, instrumental variable.

To assess the effect of FMI and FFMI on asthma (ever diagnosis and current), multivariable IVW and multivariable MR-Egger methods were fitted to the summary level data for the 60 liberal genetic variants and 39 conservative genetic variants. The BIA and DXA measurements for FMI and FFMI were considered separately. As highlighted in Section 6.2.4, the summary level data for the BIA measurements was

estimated using measurements on 360,409 participants, whereas the summary level data for the DXA measurements was only based on 3,901 of the 360,409 participants.

We refer back to the simulations performed in Sections 3.5 and 4.4 for an indication of the statistical power we may expect for the univariable and multivariable IVW analyses. For univariable IVW, we had 98.2% power (Table 3.4) to detect a positive causal effect of 0.3 when 15 genetic variants (all valid IVs) explained 3% of the variance in the risk factor with a sample size of 10,000 (see Section 3.5.1 for the data generating model). For multivariable IVW with 185 genetic variants (all valid IVs), we had 98.9% power to detect a positive causal effect of 0.3 for one risk factor when the summary level data for the three risk factors were generated independently (Table 4.5). The data generating model for this simulation study did not specify the sample size as it generated the summary level data directly.

For the multivariable MR-Egger model, the reference alleles were orientated to ensure all of the genetic associations for FMI were positive, and the estimates for FMI and the intercept term were recorded. The multivariable MR-Egger model was refitted when the genetic associations for FFMI were all positive, and the estimates for FFMI and the intercept term were recorded. These multivariable MR-Egger models were applied to the BIA and DXA measurements of FMI and FFMI. All of the Mendelian randomization analyses used random effects (see Section 2.3.1 for details).

All of the Mendelian randomization analyses were performed by Jessica Rees in RStudio version 3.5.3 [86].

Results

This Section contains the results from the one–sample Mendelian randomization study on the effect of adiposity and body composition on asthma using univariable and multivariable analyses. In Section 6.2.6, we provide a summary of the demographic characteristics of the 360,409 UK Biobank participants used to create the summary level data for the Mendelian randomization analyses. Details on the summary level data, including the R^2, adjusted R^2, and F-statistics, are contained in Section 6.2.7. The results from the univariable and multivariable Mendelian randomization analyses for an ever diagnosis of asthma are presented in Section 6.2.8. Finally, the results from the univariable and multivariable Mendelian randomization analyses for current asthma are presented in Section 6.2.9.

6.2.6 Descriptive statistics of the individual level data

A summary of the demographic characteristics of the 360,409 participants with phenotypic and QC genetic data (Figure 6.3) are recorded in Table 6.2. The mean age of the participants was 56.6 (8.0) years, and 54.2% of participants were women. 41,978 (11.6%) had reported to have been diagnosed by a doctor with asthma. Men were more likely to smoke than women, and women were less likely to do physical exercise.

Table 6.2 Summary of the demographic characteristics recorded at baseline of the 360,409 participants used to estimate the summary level data for the Mendelian randomization analyses. Mean (standard deviation) values are presented for continuous, and numbers (%) are presented for categorical measurements.

	All (N=360,409)	**Women (N=195,479)**	**Men (N=164,930)**
Age, yrs	56.6 (8.0)	56.5 (7.9)	56.9 (8.1)
BMI, kg/m^2	27.4 (4.73)	27.0 (5.12)	27.8 (4.18)
Ever diagnosis of asthma:			
Yes	41,978 (11.6)	24,088 (12.3)	17,890 (10.8)
No	318,431 (88.4)	171,391 (87.7)	147,040 (89.2)
Smoking status:			
No	323,087 (89.7)	178,244 (91.2)	144,843 (87.9)
Yes, on most days	27,100 (7.5)	12,906 (6.6)	14,194 (8.6)
Only occasionally	9,802 (2.7)	4,087 (2.0)	5,715 (3.5)
Prefer not to say	220 (0.1)	130 (0.1)	90 (0)
Missing observations	200	112	88
Alcohol consumption:			
Never	23,566 (6.5)	15,414 (7.9)	8,152 (4.9)
Special occasions only	38,239 (10.6)	27,379 (14.0)	10,860 (6.6)
1-3 times a month	40,001 (11.1)	25,479 (13.0)	14,522 (8.8)
1-2 times a week	93,973 (26.1)	51,181 (26.2)	42,792 (26.0)
3-4 times a week	86,811 (24.1)	42,254 (21.6)	44,557 (27.0)
Daily or almost daily	77,345 (21.5)	33,519 (17.2)	43,826 (26.6)
Prefer not to say	273 (0.1)	140 (0.1)	133 (0)
Missing observations	201	113	88
Walking, mins/day	54.8 (76.3)	51.5 (71.5)	58.6 (81.5)
Missing observations	13,186	6,781	6,405
Moderate activity, mins/day	59.5 (75.8)	56.0 (68.7)	63.5 (83.2)
Missing observations	60,470	33,856	26,614 (16.1)
Townsend score[a],	-1.48 (2.99)	-1.51 (2.94)	-1.44 (3.05)
Missing observations	425	231	194
Household income before tax:			
<£18,000	66,180 (18.4)	38,034 (19.5)	28,146 (17.1)
£18,000 to £30,999	78,315 (21.8)	42,585 (21.9)	35,730 (21.8)
£31,000 to £51,999	82,319 (22.9)	41,881 (21.5)	40,438 (24.7)
£52,000 to £100,000	65,933 (18.5)	31,703 (16.3)	34,230 (20.8)
>£100,000	17,738 (4.90)	8,319 (4.30)	9,419 (5.7)
Do not know	13,838 (3.80)	10,482 (5.40)	3,356 (2.0)
Prefer not to say	34,686 (9.70)	21,787 (11.1)	12,899 (7.9)
Missing observations	1,400	688	712
University or higher education[b]			
Yes	85,743 (55.3)	35,887 (42.9)	33,316 (46.7)
No	69,203 (44.7)	47,850 (57.1)	37,893 (53.3)
Missing observations	205,463	111,742	93,721

Abbreviations: yrs, years; BMI, body mass index.
[a]Higher scores represent a greater degree of deprivation.
[b]Binary variable based on the 6138 data field. Participants who stated they had a college or university degree, or other professional qualifications (e.g. nursing, teaching) were considered to have a university or higher education.

Table 6.3 contains summary measures of the adiposity and body composition measurements, and Figures I.1-I.3 contain histograms of these traits. Although there was little difference in BMI between gender, mean FMI was larger in women than men, and mean FFMI was smaller in women. The mean difference in the measurements for BIA and DXA for FMI and FFMI are also recorded in Table 6.5. On average, FMI was larger when it is measured by DXA compared to BIA, whereas FFMI was generally smaller when it is measured by BIA. There appears to be more discrepancy between the BIA and DXA measurements for men than women.

Table 6.3 Mean (standard deviation) adiposity and body composition measurements of the 360,409 participants used to estimate the summary level data for the Mendelian randomization analyses.

	All (N=360,409)	**Women (N=195,470)**	**Men (N=164,930)**
Weight, kg	78.2 (15.8)	71.4 (13.9)	86.2 (14.1)
Height, m	1.69 (0.09)	1.63 (0.06)	1.76 (0.07)
BMI, kg/m^2	27.4 (4.73)	27.0 (5.12)	27.8 (4.18)
BIA measurements, kg/m^2			
FMI	8.80 (3.61)	10.1 (3.77)	7.22 (2.65)
FFMI	18.6 (2.61)	16.8 (1.69)	20.6 (1.94)
DXA measurements, kg/m^2			
FMI	9.11 (3.41)	10.1 (3.63)	8.01 (2.77)
Missing observations	356,508	193,435	163,064
FFMI	17.2 (2.23)	15.8 (1.54)	18.8 (1.70)
Missing observations	356,508	193,435	163,064
Mean difference[a], kg/m^2			
FMI	-0.74 (1.66)	-0.4 (1.73)	-1.12 (1.5)
Missing observations	356,508	193,435	163,064
FFMI	1.22 (1.01)	0.89 (0.9)	1.58 (1.01)
Missing observations	356,508	193,435	163,064

Abbreviations: BMI, body mass index; FMI, fat mass index; FFMI, fat-free mass index; BIA, bioimpedance analysis; DXA, Dual-energy X-ray based on 3,901 participants.
[a]Mean (standard deviation) difference between bioelectrical impedance analysis and Dual-energy X-ray for fat mass index and fat-free mass index.

The mean age of asthma diagnosis was 31.1 years (Table 6.4). The mean age of diagnosis was 37.6 (20.5) years and 26.0 (15.6) years before and after the median year of birth (1950). 6,229 (17.2%) of the 41,978 participants with an ever diagnosis of asthma had been diagnosed by a doctor before the age of 9 years. The proportion of participants who reported to have a wheeze or whistle on the chest in the last year, and the mean (standard deviation) FVC and FEV1 values by an ever diagnosis of asthma are contained in Table 6.5. Using self-reported wheeze or whistling on the chest, only 27,095 (7.5%) participants were considered as current asthmatics. Of the

3,901 participants with DXA measurements, 485 (14.2%) had an ever diagnosis of asthma, and 280 (7.2%) were classed as current asthmatics.

Table 6.4 Mean (standard deviation) age of asthma diagnosis for the 41,978 who reported to have been diagnosed by a doctor with asthma by median year of birth (1950).

	All (N=41,978)	**Year of birth**	
		<1950 (N=15,620)	≥1950 (N=20,539)
Mean (SD), yrs	31.1 (18.7)	37.6 (20.5)	26.0 (15.6)
Did not know (%)	4,319 (10.7)	1,951 (13.1)	2,368 (12.0)
Preferred not to say (%)	25 (0.10)	10 (0.06)	15 (0.07)
Missing observations	1,475	676	799

Abbreviations: SD, standard deviation; yrs, years.

Table 6.5 Number (%) of participants that reported having a wheeze or whistling on the chest in the last year, and mean (standard deviation) forced vital capacity and forced expiratory volume in 1 second by ever diagnosis of asthma.

	Ever diagnosis of asthma	
	Yes (N=41,978)	No (N=325,432)
Wheeze or whistling on the chest		
Yes	27,095 (64.6)[a]	45,429 (14.0)
No	14,190 (33.8)	266,766 (82.0)
Preferred not to say	20 (0.05)	127 (0.10)
Did not know	654 (1.65)	5,927 (1.82)
Missing observations	19	182
FVC, litres	3.50 (1.05)	3.67 (1.00)
Missing observations	5,847	25,361
FEV1, litres	2.56 (0.78)	2.79 (0.77)
Missing observations	5,847	25,361

[a]These participants were classed as current asthmatics.
Abbreviations: FVC, forced volume vital capacity; FEV1, forced expiratory volume in 1 second.

6.2.7 Summary level data

Tables I.2 to I.4 and I.7 contain estimates of the genetic associations of the 60 liberal variants for BMI, FMI (BIA measurements), FFMI (BIA measurements) and asthma (ever diagnosis and current asthma) based on 360,409 participants from UK Biobank. Tables I.5 and I.6 contain estimates of the genetic associations of the DXA measurements for FMI and FFMI on 3,901 of 360,409 the participants. Two variants (rs11126666 and rs11727676) had p-values > 0.05 for both BMI and FMI (BIA), and 13 of the 60 variants did not reach GWS for BMI. None of the indices for the DXA measurements reached GWS, and a maximum of six variants were nominally associated with FMI or FFMI.

Table 6.6 contains the R^2, adjusted R^2, and F-statistics when the adiposity and body composition measurements were regressed against the liberal and conservative sets of genetic variants. The liberal variants explained 1.4% of the variance in BMI, and the conservative variants explained 0.9%. The liberal and conservative sets of genetic variants explained more variation in FMI (BIA) than FFMI (BIA). Due to the small sample size, there were noticeable differences between the R^2 and adjusted R^2 values for the DXA measurements. Both sets of genetic variants acted as weak IVs for the multivariable methods for the BIA and DXA measurements, with the largest conditional F-statistic being 5.1 for FMI (BIA).

Table 6.6 Values of the R^2, adjusted R^2, and F-statistic for the adiposity and body composition measurements from the model including the full set of liberal or conservative variants on the 360,409 participants from UK Biobank. The Sanderson-Windmeijer conditional F-statistic are provided for the BIA and DXA measurements.

	60 liberal variants				**39 conservative variants**			
	R^2	Adjusted R^2	F-stat	CF-stat	R^2	Adjusted R^2	F-stat	CF-stat
BMI	0.0137	0.0136	83.5	-	0.0093	0.0092	86.4	-
BIA measurements								
FMI	0.0094	0.0093	57.2	4.0	0.0063	0.0062	58.6	5.1
FFMI	0.0067	0.0065	40.5	3.9	0.0047	0.0046	43.4	5.0
DXA measurements[a]								
FMI	0.0170	0.0016	1.1	0.99	0.0116	0.0016	1.2	0.94
FFMI	0.0162	0.0009	1.1	0.88	0.0107	0.0007	1.1	0.74

Abbreviations: F-stat, F-statistic; CF-stat, conditional F-statistic; BMI, body mass index; BIA, bioelectrical impedance analysis; FMI, fat mass index; FFMI, fat-free mass index; DXA, Dual-energy X-ray.
[a]Estimated from the 3,901 of the 360,409 participants with DXA measurements.

The correlation between FMI and FFMI was 0.138 and 0.162 for the BIA and DXA measurements respectively. The correlation between the estimates of the genetic associations for the BIA measurements for FMI and FFMI was 0.957 and 0.952 for the liberal and conservative sets of genetic variants. For the DXA measurements, correlation between the genetic associations for FMI and FFMI was 0.658 and 0.715 for the liberal and conservative variants. Despite the values for FMI and FFMI being weakly correlated, the estimates of the genetic associations for FMI and FFMI were highly correlated for both the BIA and DXA measurements. Despite these strong correlations, we would still advise that the genetic associations for FMI and FFMI be included in the same multivariable IVW model to safe guard against the IV assumptions being violated, and to provide insight into the aetiological role of FMI and FFMI on asthma.

Figure 6.6 contains four scatter plots of the genetic associations for FMI and FFMI for the BIA and DXA measurements using the liberal and conservative sets of genetic variants selected for the Mendelian randomization analyses. Plots a. and b. in Figure 6.6 highlight the strong linear trend in the genetic associations for the BIA measurements. Only one of the genetic variants in the liberal and conservative sets of genetic variants had opposite directions of association for FMI and FFMI using the BIA measurements, and the estimate of the genetic association for FMI for this variant did not reach the nominal level of significance. Although there was a clear positive trend to the genetic associations for FMI and FFMI using the DXA measurements (plots c. and d. in Figure 6.6), only 42 of the 60 liberal genetic variants had the same direction of association for FMI and FFMI. However, the estimates for the DXA measurements were far less precise than the estimates from BIA, and few reached nominal significance for the DXA measurements.

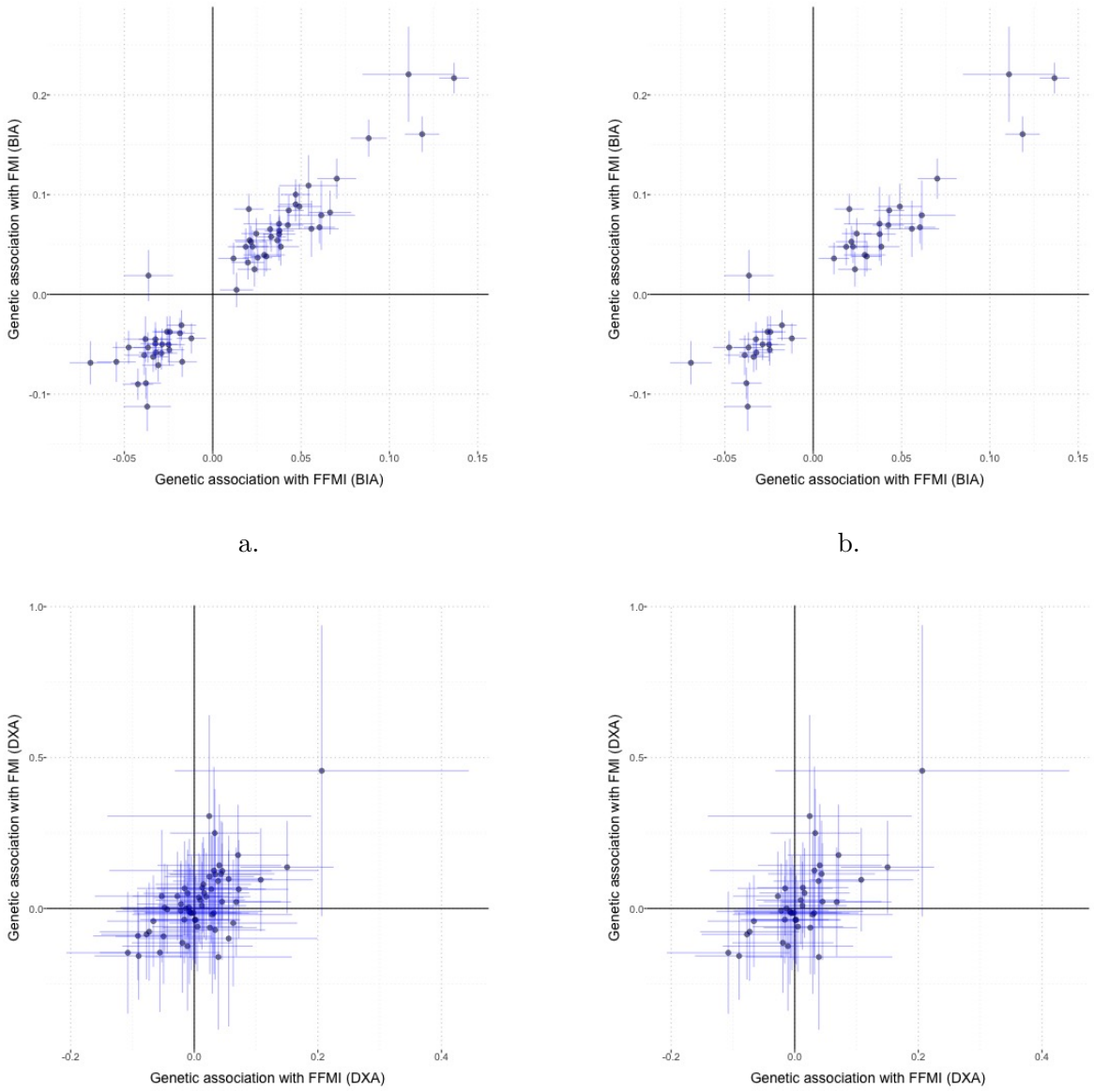


Fig. 6.6 Scatter plots of the genetic associations with 95% confidence intervals for fat mass index (FMI) and fat-free mass index (FFMI) using bioelectrical impedance analysis (BIA) for the 60 liberal (a.) and 39 conservative (b.) sets of genetic variants. The genetic associations for the BIA measurements were estimated using data on 360,409 participants from UK Biobank. The figure also contains scatter plots of the genetic associations with FMI and FFMI using dual X-ray emission absorptiometry (DXA) for the 60 liberal (c.) and 39 conservative (d.) sets of genetic variants are also displayed. The genetic associations for the DXA measurements were estimated using data on 3,901 of the 360,409 individuals from UK Biobank.

6.2.8 Ever diagnosis of asthma

In this Section, we present the results from the univariable and multivariable Mendelian randomization analyses investigating the effect of adiposity and body composition on an ever diagnosis of asthma (see Section 6.2.1 for definition).

Figure 6.7 contains the scatter plot of the genetic associations with BMI and ever diagnosis of asthma for the 60 liberal genetic variants, and Figure 6.8 contains the scatter plot of the 39 conservative genetic variants. There appears to be a positive dose-response relationship in Figures 6.7 and 6.8, although there is some heterogeneity in the genetic associations.

Results from the Mendelian randomization analyses are contained in Table 6.7. All of the estimates for BMI were positive for both sets of genetic variants, and apart from the MR-Egger model, all of the estimates were nominally associated with an ever diagnosis of asthma. There was little difference in the estimates for the liberal and conservative sets of genetic variants for the different methods. The estimate from the IVW model when the variants were treated as separate instruments was 0.047 (95% CI: 0.025, 0.069) and 0.048 (95% CI: 0.021, 0.074) for the liberal and conservative sets of genetic variants respectively. The estimate for BMI was 0.036 (95% CI: 0.002, 0.070) with a p-value of 0.038 when the two variants in the *FTO* and *MC4R* gene regions acted as IVs in the IVW model.

The estimates for FMI and FFMI using the BIA measurements from the multivariable IVW and multivariable MR-Egger models were positive and negative respectively for both sets of genetic variants. None of the estimates were significant at the nominal level, and the precision of the estimates reduced when the conservative set of variants were used as IVs in the multivariable IVW model.

The estimates for FMI using the DXA measurements were both positive from the multivariable IVW and multivariable MR-Egger models for the two sets of variants. The estimates for FMI using the DXA measurements reached nominal significance in the multivariable IVW model for the liberal (0.041, 95% CI: 0.001, 0.081) and conservative (0.050, 95% CI: 0.004, 0.095) sets of genetic variants. Although the estimates for FMI (DXA) from the multivariable MR-Egger model did reach nominal significance for the two sets of genetic variants, the p-value of the intercept term was 0.008 and 0.023 for the liberal and conservative variants respectively.

Fig. 6.7 Scatter plot of the genetic associations with body mass index and ever diagnosis of asthma (41,978 cases) for the 60 liberal variants included in the Mendelian randomization analyses. The genetic associations were estimated using data on 360,409 participants from UK Biobank. The blue line represents the estimate of the approximate causal log odds ratio for an ever diagnosis of asthma per unit increase in body mass index from the IVW method.

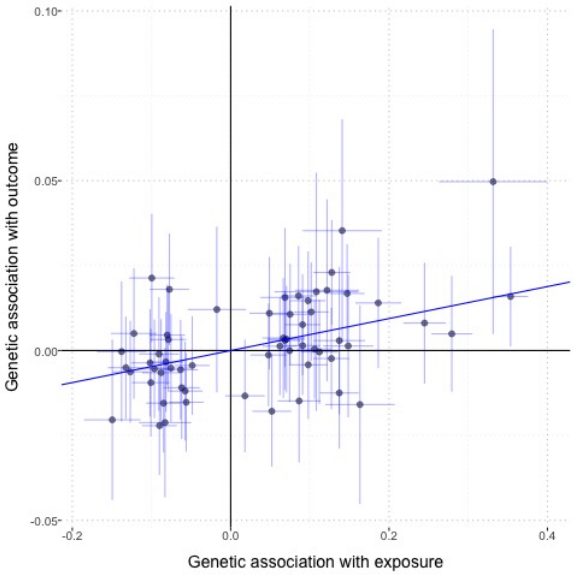


Fig. 6.8 Scatter plot of the genetic associations with body mass index and ever dignosis of asthma (41,978 cases) for the 39 conservative variants included in the Mendelian randomization analyses. The genetic associations were estimated using data on 360,409 participants from UK Biobank. The blue line represents the estimate of the approximate causal log odds ratio for an ever diagnosis of asthma per unit increase in body mass index from the IVW method.

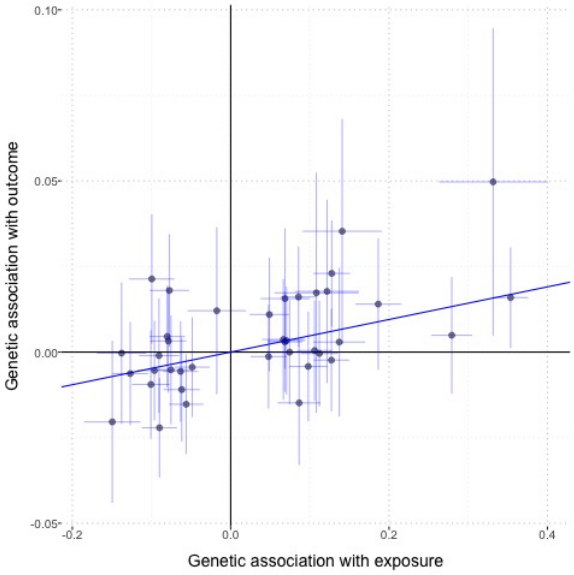


Table 6.7 Estimates of the approximate causal log odds ratios (standard errors) for ever diagnosis of asthma per unit increase in body mass index from IVW, simple median, weighted median and MR-Egger (with intercept estimate) methods. The table contains estimates of the approximate causal log odds ratios for ever diagnosis of asthma per unit increase in the BIA and DXA measurements from multivariable IVW and multivariable MR-Egger (with intercept estimate). Estimates from multivariable MR-Egger are presented for each body composition measurement where the reference allele is the trait increasing allele for the measurement. Estimates are provided for the liberal and conservative sets of genetic variants.

	60 liberal variants			**39 conservative variants**		
	Estimate (se)	95% CI	P-value	Estimate (se)	95% CI	P-value
BMI						
IVW	0.047 (0.011)	0.025, 0.069	2×10^{-5}	0.048 (0.014)	0.021, 0.074	4×10^{-4}
Median	0.046 (0.016)	0.016, 0.077	0.003	0.050 (0.020)	0.010, 0.089	0.014
Weighted median	0.044 (0.016)	0.014, 0.075	0.004	0.045 (0.018)	0.010, 0.080	0.012
MR-Egger	0.045 (0.023)	0.000, 0.090	0.050	0.046 (0.026)	-0.004, 0.097	0.074
(Intercept)	0.000 (0.003)	-0.005, 0.006	0.904	0.000 (0.003)	-0.006, 0.006	0.949
BIA measurements						
Multivariable IVW						
FMI	0.119 (0.066)	-0.010, 0.248	0.070	0.077 (0.079)	-0.077, 0.232	0.328
FFMI	-0.075 (0.111)	-0.292, 0.142	0.497	-0.001 (0.128)	-0.253, 0.251	0.994
Multivariable MR-Egger						
FMI	0.113 (0.076)	-0.039, 0.266	0.141	0.062 (0.094)	-0.128, 0.252	0.512
(Intercept)	0.000 (0.003)	-0.005, 0.006	0.881	0.001 (0.003)	-0.006, 0.008	0.758
FFMI	-0.075 (0.112)	-0.299, 0.149	0.505	-0.001 (0.130)	-0.265, 0.263	0.994
(Intercept)	0.000 (0.003)	-0.005, 0.005	0.982	0.000 (0.003)	-0.007, 0.007	0.994
DXA measurements[a]						
Multivariable IVW						
FMI	0.041 (0.021)	0.001, 0.081	0.046	0.050 (0.023)	0.004, 0.095	0.033
FFMI	0.009 (0.038)	-0.065, 0.083	0.815	0.021 (0.044)	-0.065, 0.107	0.631
Multivariable MR-Egger						
FMI	0.092 (0.026)	0.039, 0.145	0.001	0.091 (0.028)	0.035, 0.147	0.002
(Intercept)	-0.006 (0.002)	-0.010, -0.002	0.006	-0.006 (0.002)	-0.011, -0.001	0.019
FFMI	0.045 (0.050)	-0.055, 0.145	0.371	0.052 (0.055)	-0.059, 0.162	0.349
(Intercept)	-0.002 (0.002)	-0.007, 0.002	0.274	-0.002 (0.002)	-0.007, 0.003	0.349

Abbreviations: se, standard error; CI, confidence interval; BMI, body mass index; IVW, inverse-variance weighted; FMI, fat mass index; FFMI, fat-free mass index; BIA, bioelectrical impedance analysis; DXA, Dual-energy X-ray.
[a]Results based on the genetic associations for the DXA measurements estimated on 3,901 of the 360,409 participants, and the genetic associations for ever diagnosis of asthma estimated on 360,409 participants.

6.2.9 Current asthma

In this Section, we present the results from the univariable and multivariable Mendelian randomization analyses investigating the effect of adiposity and body composition on current asthma (see Section 6.2.1 for definition).

Figures 6.9 and 6.10 contain the scatter plots of the genetic associations with BMI and current asthma status for the liberal and conservative sets of genetic variants. There is little difference between Figures 6.9 and 6.10, and we still observe positive dose-response relationships when current asthma is considered rather than an ever diagnosis of asthma (Figure 6.7 and 6.8).

The results from the Mendelian randomization models for current asthma are contained in Table 6.8. The estimates for BMI increased for all models compared to the estimates in Table 6.7. The precision of the estimates decreased when current asthma (Table 6.8) was considered rather than an ever diagnosis of asthma (Table 6.7). The estimate from the IVW model with the two variants from the *FTO* and *MC4R* gene region was 0.059 (95% CI: 0.017, 0.100) with p-value of 0.006. The estimates from the MR-Egger model when the liberal (0.056, 95% CI: 0.003, 0.109) and conservative (0.071, 95% CI: 0.012, 0.129) sets of variants acted as IVs were significant at the nominal level.

The results from the multivariable models were similar to an ever diagnosis of asthma (Table 6.7). The estimates for FMI (DXA) still reached the nominal level of significance for the multivariable IVW and multivariable MR-Egger models. The estimates of the intercept term from the multivariable MR-Egger models had a p-value < 0.05 for both sets of variants.

Fig. 6.9 Scatter plot of the genetic associations with body mass index and current asthma status (27,095 cases) with 95% confidence intervals for the 60 liberal variants included in the Mendelian randomization analyses. The genetic associations were estimated using data on 360,409 participants from UK Biobank. The blue line represents the estimate of the approximate causal log odds ratio for current diagnosis of asthma per unit increase in body mass index from the IVW method.

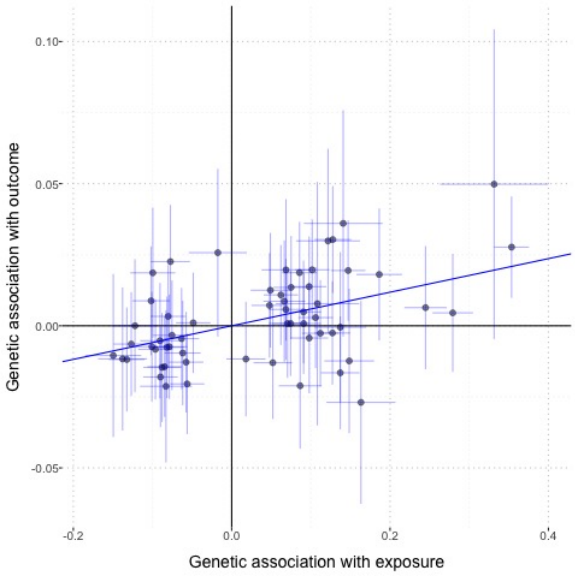


Fig. 6.10 Scatter plot of the genetic associations with body mass index and current asthma status (27,095 cases) with 95% confidence intervals for the 39 conservative variants included in the Mendelian randomization analyses. The genetic associations were estimated using data on 360,409 participants from UK Biobank. The blue line represents the estimate of the approximate causal log odds ratio for current diagnosis of asthma per unit increase in body mass index from the IVW method.

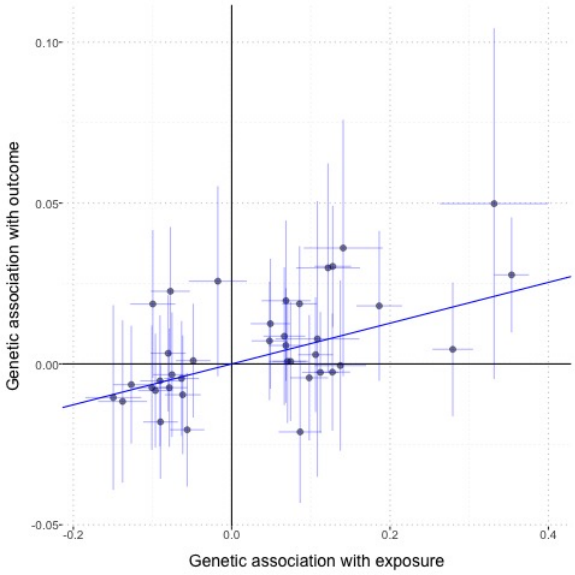


Table 6.8 Estimates of the approximate causal log odds ratios (standard errors) for current asthma per unit increase in body mass index from IVW, simple median, weighted median and MR-Egger (with intercept estimate) methods. The table contains the estimates of the approximate causal log odds ratios for current asthma per unit increase in the BIA and DXA measurements from multivariable IVW and multivariable MR-Egger (with intercept estimate). Estimates from multivariable MR-Egger are presented for each body composition measurement where the reference allele is the trait increasing allele for the measurement. Estimates are provided for the liberal and conservative sets of genetic variants.

	60 liberal variants			**39 conservative variants**		
	Estimate (se)	95% CI	P-value	Estimate (se)	95% CI	P-value
BMI						
IVW	0.059 (0.013)	0.033, 0.085	1×10^{-5}	0.063 (0.016)	0.033, 0.094	5×10^{-5}
Simple median	0.072 (0.019)	0.035, 0.110	2×10^{-4}	0.072 (0.024)	0.025, 0.119	0.003
Weighted median	0.075 (0.018)	0.039, 0.110	5×10^{-5}	0.075 (0.022)	0.033, 0.117	0.001
MR-Egger	0.056 (0.027)	0.003, 0.109	0.039	0.071 (0.030)	0.012, 0.129	0.018
(Intercept)	0.000 (0.003)	-0.006, 0.007	0.892	-0.001 (0.004)	-0.008, 0.006	0.778
BIA measurements						
Multivariable IVW						
FMI	0.132 (0.078)	-0.020, 0.285	0.090	0.107 (0.091)	-0.071, 0.284	0.241
FFMI	-0.065 (0.131)	-0.323, 0.192	0.619	-0.007 (0.148)	-0.297, 0.282	0.960
Multivariable MR-Egger						
FMI	0.112 (0.090)	-0.068, 0.293	0.217	0.094 (0.108)	-0.125, 0.313	0.390
(Intercept)	0.001 (0.003)	-0.005, 0.008	0.658	0.001 (0.004)	-0.007, 0.009	0.826
FFMI	-0.065 (0.133)	-0.331, 0.200	0.624	-0.006 (0.150)	-0.310, 0.297	0.968
(Intercept)	0.000 (0.003)	-0.006, 0.006	0.989	-0.001 (0.004)	-0.009, 0.006	0.726
DXA measurements[a]						
Multivariable IVW						
FMI	0.055 (0.024)	0.007, 0.102	0.025	0.063 (0.027)	0.010, 0.116	0.019
FFMI	0.007 (0.045)	-0.080, 0.095	0.869	0.034 (0.051)	-0.066, 0.134	0.503
Multivariable MR-Egger						
FMI	0.110 (0.032)	0.046, 0.173	0.001	0.111 (0.032)	0.046, 0.176	0.001
(Intercept)	-0.006 (0.002)	-0.011, -0.001	0.013	-0.007 (0.003)	-0.012, -0.001	0.021
FFMI	0.045 (0.059)	-0.074, 0.163	0.455	0.045 (0.064)	-0.085, 0.175	0.485
(Intercept)	-0.002 (0.003)	-0.008, 0.003	0.343	-0.001 (0.003)	-0.006, 0.005	0.773

Abbreviations: se, standard error; CI, confidence interval; BMI, body mass index; IVW, inverse-variance weighted; FMI, fat mass index; FFMI, fat-free mass index; BIA, bioelectrical impedance analysis; DXA, Dual-energy X-ray.
[a]Results based on the genetic associations for the DXA measurements estimated on 3,901 of the 360,409 participants, and the genetic associations for ever diagnosis of asthma estimated on 360,409 participants.

6.2.10 Summary

In this Section, we have performed a one-sample Mendelian randomization analysis to investigate the effect of adiposity and body composition on an ever diagnosis of asthma and current asthma. We have presented evidence of a positive causal association between BMI and asthma for both an ever diagnosis of asthma and current asthma for all sets of IVs considered in the IVW models. The effect of body composition (measured through FMI and FFMI) on asthma was less clear. None of the measurements for FMI and FFMI under BIA reached the nominal level of significance for the multivariable analyses, although the direction of the point estimates suggested that FMI may have an adverse effect on asthma. This was supported from the estimates for the DXA measurements where FMI was nominally associated with asthma in all of the multivariable models. However, the intercept term from the multivariable MR-Egger model for FMI was significant for all models, suggesting that the InSIDE assumption for multivariable MR-Egger was not satisfied and/or there was unmeasured pleiotropy [32].

6.3 Two–sample Mendelian randomization analysis on the effects of adiposity and body composition on asthma

In this Section, we perform a two–sample Mendelian randomization analysis using summary level data. For this Mendelian randomization study, summary level data is estimated under the conventional setting (Section 1.5.2): the genetic associations with the risk factor and the genetic associations with the outcome are estimated in two separate samples. The genetic associations with BMI, FMI (BIA and DXA measurements) and FFMI (BIA and DXA measurements) are estimated from UK Biobank (predominantly taken from Section 6.2), and the estimates of the genetic associations with asthma are taken from the GABRIEL Consortium. The GABRIEL consortium consists of 10,365 asthma cases (diagnosed by a doctor) and 16,110 controls from 23 studies. All of the studies provided data on individuals diagnosed with asthma before the age of 16 years, and over half of the studies provided data on adults (diagnosed with asthma 16 years or older). A GWAS on asthma was performed on the entire dataset [37].

This two–sample Mendelian randomization study used the same methods considered for the one–sample Mendelian randomization study (Section 6.2). To allow for comparisons between the one–sample and two–sample estimates, the genetic variants selected under the liberal and conservative criteria in Section 6.2.3 were considered for this two–sample Mendelian randomization study. Section 6.3.1 provides an overview of the summary level data and the methods used in the analysis, and the results are presented in Section 6.3.2.

6.3.1 Methods

Figure 6.11 provides an overview of the summary level data used in the Mendelian randomization analyses. In the first step, the rs numbers of the 60 genetic variants selected under the liberal criteria in Section 6.2.3 were searched in Phenoscanner by Jessica Rees with the option of searching for genetic variants that could act as proxies for the 60 genetic variants (no cut-off value was used for the p-value of the genetic associations). Phenoscanner identified proxies of the 60 genetic variants by using a pairwise r^2 (based on the 1,000 Genome Project) threshold of 0.8. Hence, a genetic variant was classed as a proxy variant if its r^2 value with one of the specified 60 genetic variants was greater than 0.8. As noted in Section 6.2.3, r^2 takes into account the

linkage disequilibrium between the genetic variants and the minor allele frequency of the genetic variants.

The genetic associations with asthma from the GABRIEL consortium [37] were then extracted from the Phenoscanner results obtained from the above search. Since the option for searching for proxy variants had been used, some of the genetic variants had more than one estimate from the GABRIEL consortium. The following selection criteria was used to ensure that each of the 60 liberal variants only had one estimate for the genetic association with asthma: 1) if there was an estimate of the genetic association on the exact variant then this was retained; and 2) if there was no estimate for the exact variant but there were estimates for proxy variants, then the genetic association estimate with the proxy variant with the highest r^2 value was retained.

42 of the 60 genetic variants searched in Phenoscanner had genetic associations (either with the exact variant or a proxy variant) with asthma in the GABRIEL consortium (Figure 6.11). Of these 42 genetic variants, 23 had estimates based on the exact variant, and 19 had estimates based on a proxy variant. 29 of the 39 genetic variants selected under the conservative criteria in Section 6.2.3 had genetic association estimates for asthma from the GABRIEL consortium. Of the 29 genetic variants in the conservative set, 16 had estimates based on the exact variant, and 13 had estimates based on a proxy variant. Table I.8 provides the rs number and r^2 value of the proxy variant when relevant, and contains information on whether the variant was included in the conservative set of variants.

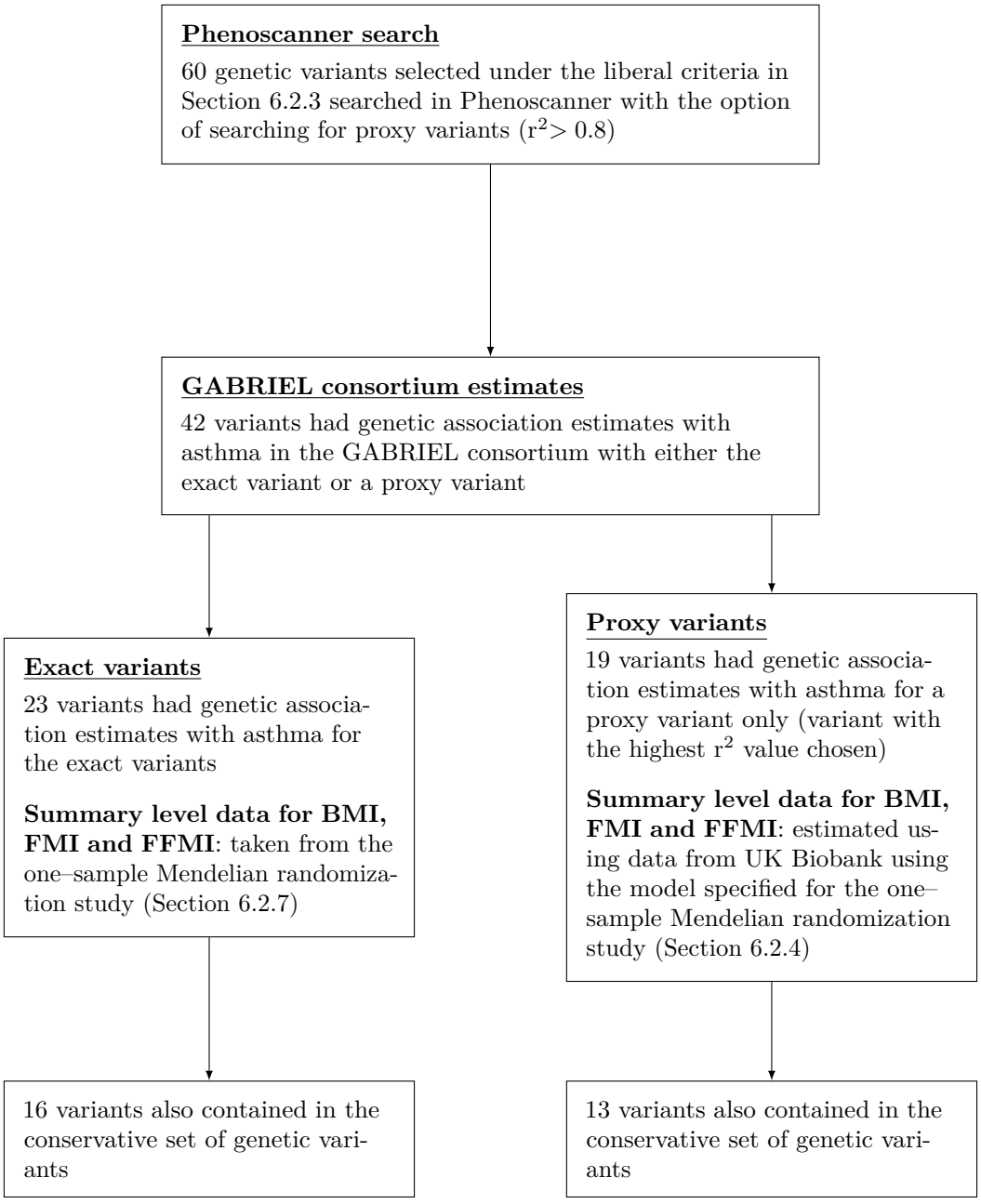
As highlighted in Figure 6.11, the genetic associations with BMI, FMI (BIA and DXA measurements) and FFMI (BIA and DXA measurements) were taken from the one–sample Mendelian randomization study (Section 6.2.4) for the 23 variants that had genetic association estimates for asthma based on the exact variant. The genetic associations with BMI, FMI (BIA and DXA measurements) and FFMI (BIA and DXA measurements) were re-estimated using the UK Biobank dataset for the 19 variants whose genetic association estimate for asthma was based on a proxy variant. Jessica Rees used the *QCTOOL v2* command–line programme to extract the imputed genetic data on the 19 proxy variants for the 360,409 participants from UK Biobank. As outlined in Section 6.2.4, *SNPTEST v2* was used to estimate the genetic associations for BMI, FMI (BIA and DXA measurements) and FFMI (BIA and DXA measurements) adjusted for the first 10 ancestry informative PCs, gender and height.

The genetic associations with BMI from UK Biobank were plotted against the genetic associations with asthma from the GABRIEL consortium for the 42 liberal and

29 conservative sets of genetic variants. These plots were created by Jessica Rees in RStudio version 3.5.3 [86] using the package *ggplot2* [141].

With the exception of the IVW method with the two genetic variants from the gene regions *FTO* and *MC4R*, all of the univariable and multivariable Mendelian randomization analyses outlined in Section 6.2.5 were performed on the summary level data for the 42 liberal and 29 conservative sets of genetic variants. All of the analyses were performed by Jessica Rees in RStudio version 3.5.3 [86].

Fig. 6.11 Flowchart highlighting the number of variants with genetic association estimates with asthma in the GABRIEL consortium by whether the estimates are based on the exact or proxy variant. The source of the summary level data for BMI, FMI and FFMI is also outlined.



Abbreviations: BMI, body mass index; FMI, body fat mass index; and FFMI, body fat-free mass index.

6.3.2 Results

The genetic associations of the 19 proxy variants with BMI, FMI (BIA and DXA measurements) and FFMI (BIA and DXA measurements) from UK Biobank are displayed in Table I.9. 3 proxy variants for BMI, 5 for FMI (BIA), and 4 for FFMI (BIA) did not reach GWS significance. Figures 6.12 and 6.13 contain the scatter plots of the genetic associations of BMI against the genetic associations of asthma. There appeared to be no obvious trend in the associations, and there was clear heterogeneity in the approximate causal estimates.

The results from the Mendelian randomization analyses are contained in Table 6.9. The IVW estimates for the liberal variants and conservative variants were 0.040 (95% CI: -0.030, 0.110) and 0.050 (95% CI: -0.034, 0.134) respectively. Although the IVW estimates for BMI were positive, they did not reach nominal significance as seen in the one-sample setting (Section 6.2). The median estimators and MR-Egger model did not provide any evidence of a causal effect of BMI on asthma.

As seen in the one-sample setting (Section 6.2), the multivariable IVW and multivariable MR-Egger models for the BIA measurements did not provide any evidence of a causal effect of FMI or FFMI on asthma for both sets of genetic variants. Apart from the estimates from the multivariable IVW model with the liberal variants, the point estimates for FMI (BIA) and FFMI (BIA) were positive and negative respectively for the multivariable IVW and multivariable MR-Egger models. The intercept term for the liberal and conservative sets of variants were not significant at the nominal level.

The point estimates for FMI and FFMI using the DXA measurements were positive and negative respectively for all of the multivariable models. The estimates for FMI (DXA) from the multivariable IVW method were borderline significant for the liberal (0.112, 95% CI: -0.002, 0.225) and conservative (0.137, 95% CI: -0.004, 0.278) sets of genetic variants. There was no evidence of a causal effect of FFMI (DXA) on asthma. Unlike Section 6.2, the intercept term for the multivariable MR-Egger model did not reach nominal significance for either the liberal or conservative sets of variants.

Fig. 6.12 Scatter plot of the genetic associations with body mass index from UK Biobank and the genetic associations with asthma from the GABRIEL consortium (10,365 physician diagnosed cases) for the 42 liberal variants (19 were proxy variants) with 95% confidence intervals. The genetic associations for BMI were estimated using data on 360,409 participants from UK Biobank. The blue line represents the estimate of the approximate causal log odds ratio for asthma per unit increase in body mass index from the IVW model.

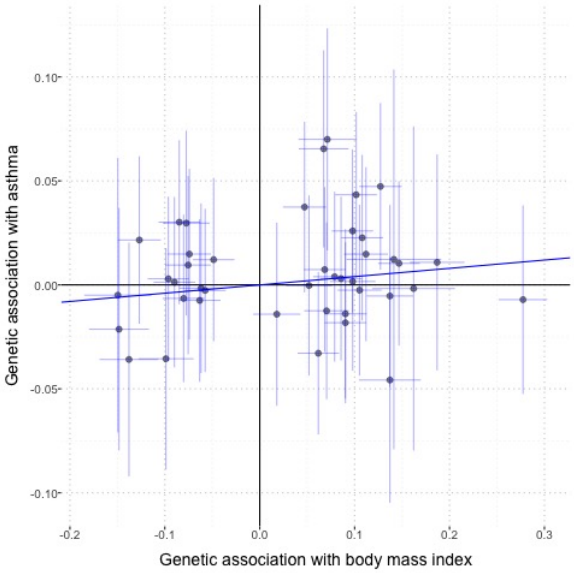


Fig. 6.13 Scatter plot of the genetic associations with body mass index from UK Biobank and the genetic associations with asthma from the GABRIEL consortium (10,365 physician diagnosed cases) for the 29 conservative variants (13 were proxy variants) with 95% confidence intervals. The genetic associations for BMI were estimated using data on 360,409 participants from UK Biobank. The blue line represents the estimate of the approximate causal log odds ratio for asthma per unit increase in body mass index from the IVW model.

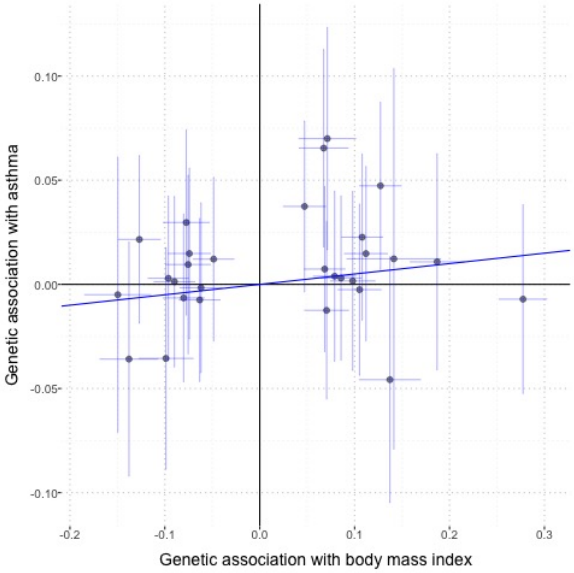


Table 6.9 Estimates of the approximate causal log odds ratios (standard errors) for asthma per unit increase in body mass index from IVW, simple median, weighted median and MR-Egger (with intercept estimate) methods for the liberal and conservative sets of genetic variants. The table also contains the estimate of the approximate causal log odds ratios for asthma per unit increase in the BIA and DXA measurements from multivariable IVW and multivariable MR-Egger (with intercept estimate).

	42 liberal variants[a]			**29 conservative variants**[b]		
	Estimate (se)	95% CI	P-value	Estimate (se)	95% CI	P-value
IVW	0.040 (0.036)	-0.030, 0.110	0.261	0.050 (0.042)	-0.034, 0.134	0.242
Simple median	0.030 (0.051)	-0.070, 0.129	0.556	0.035 (0.061)	-0.085, 0.156	0.566
Weighted median	0.010 (0.050)	-0.088, 0.108	0.839	0.005 (0.058)	-0.109, 0.119	0.931
MR-Egger	0.031 (0.088)	-0.142, 0.204	0.726	-0.043 (0.102)	-0.244, 0.158	0.674
(Intercept)	0.001 (0.009)	-0.017, 0.019	0.911	0.011(0.011)	-0.010, 0.032	0.318
BIA measurements						
Multivariable IVW						
FMI	0.055 (0.200)	-0.338, 0.448	0.783	0.157 (0.238)	-0.311, 0.623	0.511
FFMI	0.013 (0.346)	-0.665, 0.692	0.969	-0.132 (0.401)	-0.919, 0.656	0.743
Multivariable MR-Egger[c]						
FMI	0.053 (0.197)	-0.342, 0.447	0.791	0.136 (0.228)	-0.327, 0.599	0.556
FFMI	-0.026 (0.341)	-0.710, 0.658	0.940	-0.132 (0.384)	-0.911, 0.647	0.733
(Intercept)	0.006 (0.004)	-0.002, 0.014	0.129	0.008 (0.004)	-0.001, 0.017	0.071
DXA measurements[d]						
Multivariable IVW						
FMI	0.112 (0.058)	-0.002, 0.225	0.054	0.137 (0.072)	-0.004, 0.278	0.057
FFMI	-0.087 (0.100)	-0.284, 0.109	0.383	-0.114 (0.127)	-0.362, 0.134	0.369
Multivariable MR-Egger						
FMI	0.166 (0.090)	-0.015, 0.346	0.074	0.184 (0.106)	-0.031, 0.400	0.095
(Intercept)	-0.005 (0.006)	-0.017, 0.007	0.441	-0.005 (0.007)	-0.020, 0.011	0.548
FFMI	-0.229 (0.147)	-0.522, 0.064	0.126	-0.305 (0.163)	-0.636, 0.026	0.073
(Intercept)	0.007 (0.006)	-0.004, 0.019	0.196	0.011 (0.006)	-0.002, 0.024	0.090

Abbreviations: se, standard error; CI, confidence interval; BMI, body mass index; IVW, inverse-variance weighted; FMI, fat mass index; FFMI, fat-free mass index; BIA, bioelectrical impedance analysis; DXA, Dual-energy X-ray.
[a]19 of the 42 genetic variants were proxy variants. [b]13 of the 29 genetic variants were proxy variants. [c]Multivariable MR-Egger model only fitted once as the direction of association for FMI and FFMI was the same across the 42 genetic variants.
[d]Results based on data from 3,901 of the 360,409 participants.

6.3.3 Summary

In this Section, we have performed a two-sample Mendelian randomization analysis using summary level data from UK Biobank and the GABRIEL consortium. Contrary to the results obtained in Section 6.2, we have found no evidence of a causal association between BMI and asthma. The multivariable methods supported the results in Section 6.2 of a null causal effect of FMI and FFMI on asthma for the BIA measurements. The estimate for FMI using the DXA measurements were borderline significant in the

multivariable IVW model. In the next Section, we compare the results from the Mendelian randomization studies that have been performed in the Chapter.

6.4 Comparison of the results from the Mendelian randomization analyses

In this Section, we compare the results from the Mendelian randomization analyses performed in Sections 6.2 and 6.3.

6.4.1 Univariable analyses

Figure 6.14 contains the log odds ratios with 95% confidence intervals for asthma per unit increase in BMI from the IVW models when the liberal (green bars) and conservative (blue bars) sets of genetic variants were used as IVs in Sections 6.2 and 6.3. For the one–sample Mendelian randomization analysis, the IVW estimates are based on an ever diagnosis of asthma (as contained in Section 6.2.8). There were 60 liberal and 39 conservative genetic variants for the one–sample Mendelian randomization analysis. The results from the two–sample Mendelian randomization analyses are based on 42 liberal (19 were proxy variants) and 29 conservative (13 were proxy variants) genetic variants.

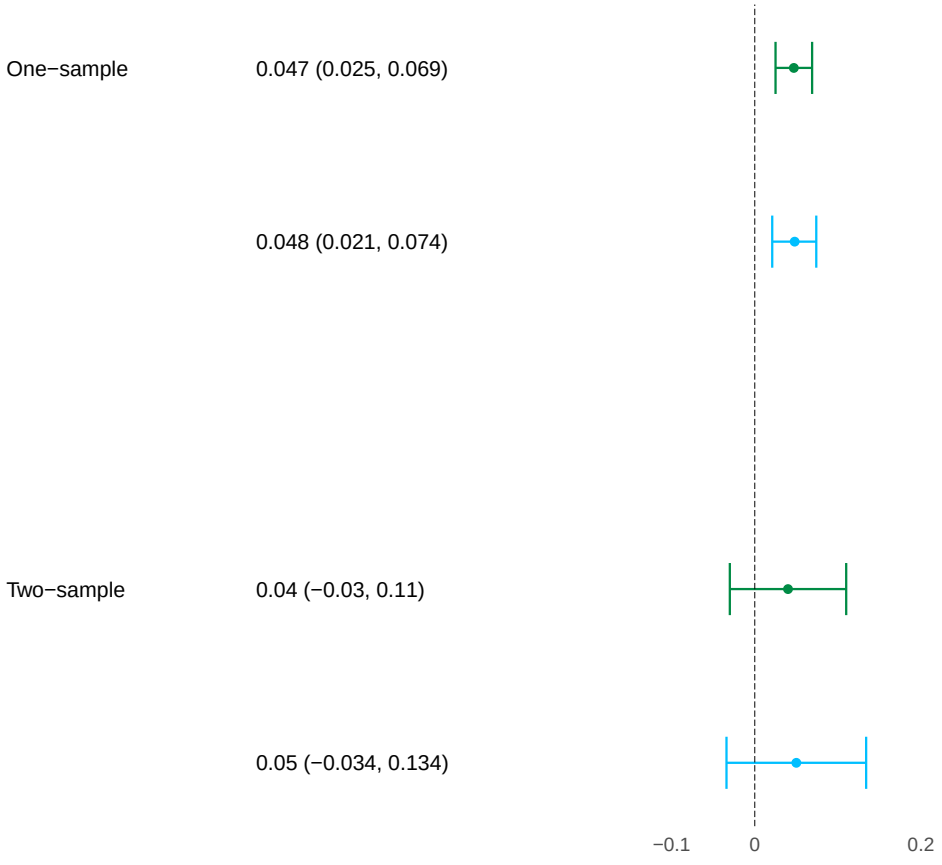
Both estimates from the one–sample Mendelian randomization study were nominally significant, and the magnitude of the estimates were very similar. The point estimates from the two-sample Mendelian randomization study were similar to the estimates obtained from the one–sample. Unlike the one–sample estimates, the two-sample estimates did not reach nominal significance. This discrepancy in the results may be due to the GABRIEL consortium containing data on children and adults, whereas the one–sample analysis was only based on UK Biobank data where participants were adults aged between 40-69 years. There were also differences in the genetic variants used as IVs for the one–sample and two-sample studies. Although all of the 60 liberal genetic variants were considered for the two-sample study, we could only obtain estimates of the genetic associations with asthma on 42 of the 60 variants, and 19 of these variants were proxies.

The discrepancy between the results from the one–sample and the two-sample studies may have also arisen by the GABRIEL consortium requiring a physician diagnosis for the participant to be considered an asthma case, whereas we used self-reported measurements to define asthma status in UK Biobank. In fact, the summary level data from the GABRIEL consortium was based on data from 26,475 participants, with 10,365 (39.2%) being classed as asthma cases. The genetic associations for asthma from UK Biobank were based on 360,409 participants, with 41,978 (11.6%) classed as

having an ever diagnosis of asthma. This difference in the proportion of participants classed as asthmatic highlights the differences in the study designs for UK Biobank and the GABRIEL consortium.

For the one–sample study, we estimated the summary level data for the risk factors and the outcome on the same set of individuals. As highlighted in Section 6.1.4, we apply this summary level data to methods specifically designed for the two–sample setting (IVW, median and MR-Egger). These methods assume that the genetic associations for the risk factor and outcome are estimated from two unrelated samples. For the two–sample case, the genetic associations were estimated from two samples. This may have contributed to the discrepancy between the estimates obtained from the one–sample and two–sample setting.

Fig. 6.14 Estimates of the approximate causal log odds ratios with 95% confidence intervals for asthma per unit increase in body mass index from the IVW models for the one–sample (estimates from an ever diagnosis of asthma) and two–sample Mendelian randomization studies. The green bars represent the estimates from the liberal sets of genetic variants, and the blue bars represent the estimates from the conservative sets of genetic variants.



6.4.2 Multivariable analyses

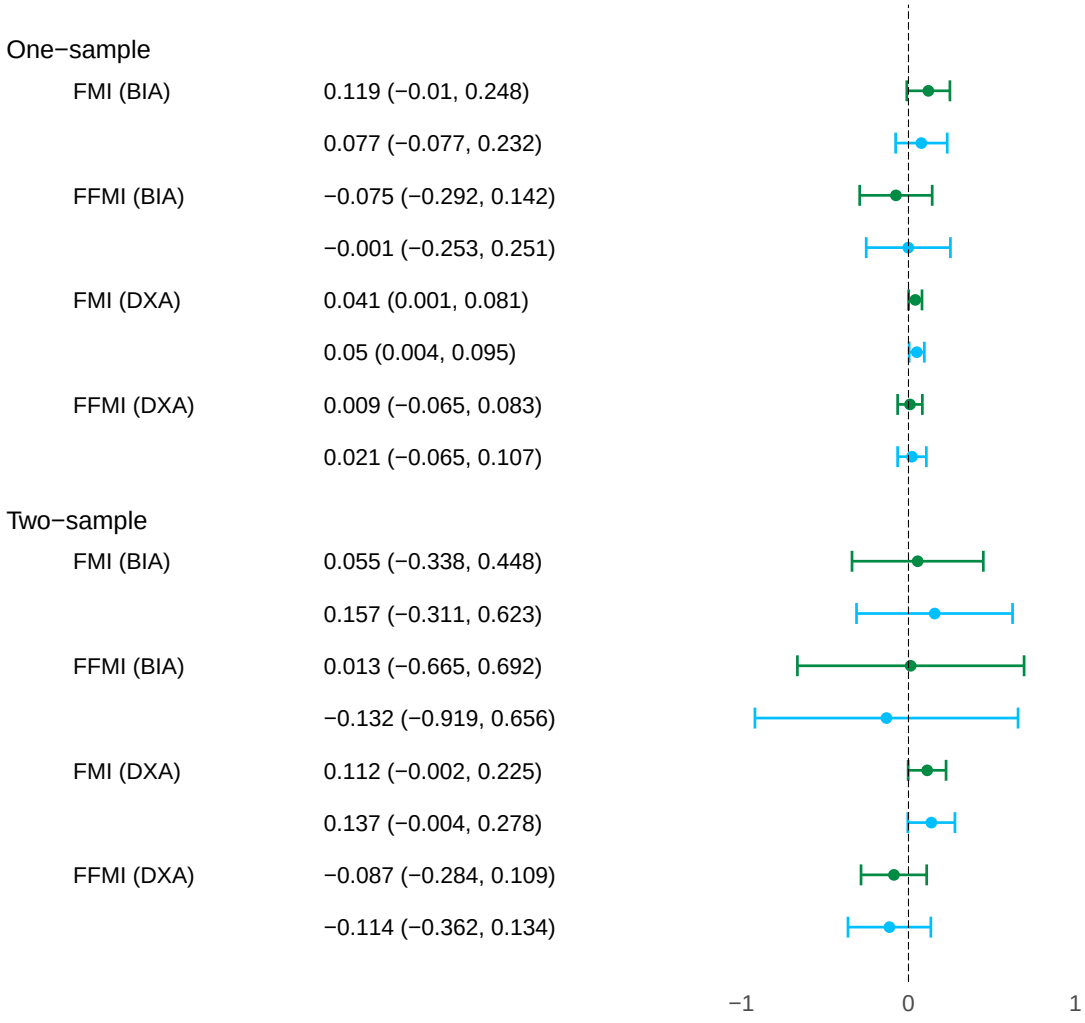
Figure 6.15 contains the log odds ratios with 95% confidence intervals for asthma per unit increase in FMI and FFMI from the multivariable IVW models for the BIA and DXA measurements from Sections 6.2 and 6.3.

Figure 6.15 highlights the differences in the precision of the estimates for the BIA and DXA measurements. The estimates for FMI and FFMI from the one–sample and two–sample analyses were more precise for the DXA measurements than the BIA measurements. This result seems somewhat surprising considering the genetic associations for the DXA measurements were only based on 3,901 of the 360,409 participants. The differences in the precision between the BIA and DXA measurements may be due to: 1) the DXA measurements capturing more variability in the FM and FFM measurements; 2) the FMI and FFMI genetic associations being less correlated under the DXA measurements; or 3) a systematic difference in the characteristics of the participants in the complete dataset (N=360,409) compared to the participants with BIA and DXA measurements (N=3,901).

The results obtained from the multivariable methods should be interpreted with some caution, particularly the positive results for the DXA measurement of FMI as there was evidence of weak instruments. As discussed in Section 2.6.3, a genetic variant is classed as a strong instrument in the multivariable setting if: a) it is strongly associated with all of the explanatory variables (measured through the F-statistic); and b) it is jointly associated with the explanatory variables (measured through the conditional F-statistic). Although some of the sets of genetic variants were strongly associated with the BIA measurements, the conditional F-statistics were small for all of the BIA and DXA measurements (Section 6.2.4). Since UK Biobank will be releasing additional data on the DXA measurements, it would be advisable for the multivariable models to be refitted to the larger dataset to increase the strength of the IVs.

The results from the two-sample analysis supported the null causal effect of FMI on asthma in the analysis using BIA on the UK Biobank dataset. Unlike the estimates from the one-sample study, the two-sample study provided evidence of a positive causal effect of FMI on asthma using the DXA measurements. Possible reasons for the discrepancy between the results obtained from the one–sample and two–sample analyses have been highlighted for the univariable case (Section 6.4.1), and these may also be applicable to the multivariable case.

Fig. 6.15 Estimate of the approximate causal log odds ratios with 95% confidence intervals for asthma per unit increase in fat mass index and fat-free mass index from the multivariable IVW models for the one–sample (estimates from an ever diagnosis of asthma) and two-sample Mendelian randomization studies. The green bars represent the estimates from the liberal sets of genetic variants, and the blue bars represent the estimates from the conservative sets of genetic variants.



Abbreviations: FMI, fat mass index; FFMI, fat-free mass index; LMI, lean mass index; BIA, bioelectrical impedance analysis; DXA, Dual-energy X-ray.

6.4.3 Summary

In this Section, we have compared the estimates from the univariable and multivariable analyses for the one–sample and two–sample Mendelian randomization studies. We have highlighted the similarities in the point estimates obtained from the one–sample and two–sample Mendelian randomization studies, and have discussed possible reasons for the null results in the two-sample study. The estimates from the multivariable IVW models have also been compared. Potential issues with weak instrument bias in the multivariable models have been highlighted, particularly in relation to the positive results for the DXA measurement of FMI. Differences in the precision of the estimates for the BIA and DXA measurements have also been highlighted.

6.5 Discussion

In this Chapter, we have performed univariable and multivariable Mendelian randomization analyses in the one–sample and two-sample setting to investigate the effect of adiposity and body composition measurements on asthma. The one-sample Mendelian randomization analyses in UK Biobank supported the epidemiological evidence of there being a positive causal association between BMI and asthma in adults, yet the two-sample study produced a null result. The multivariable models for the DXA measurements suggested that increased FMI had a positive causal effect on asthma, but this result was not replicated in the BIA measurements.

6.5.1 Measurements for asthma and body composition

Since asthma status in UK Biobank was based on self-report, the possibility of introducing bias into the analysis through misclassification of the outcome is a concern. There has not been any validation studies on asthma status in UK Biobank, and this should be explored further, especially since there are no questions that directly relate to current asthma status. In fact, since asthma is usually developed during childhood, and it is possible to outgrow the condition, some of the participants in UK Biobank who were classed as asthmatic under the ever diagnosis criteria may not have been asthmatic at recruitment. We hoped to have a better appreciation for the number of participants who were current asthmatics by including wheezing or whistling on the chest in our second measure for asthma status.

It is often assumed that BMI is fully adjusted for height. Since body composition varies with age and gender, this assumption is not always true, and rather than squaring height, it may be more appropriate to apply a different power (BMI[x]=weight(kg)/height(m^x)) [194]. This logic could also be applied to the other index measurements of body composition. To help with interpretation and comparison between the measurements, we decided that all of the indices should be calculated using the square of height. We acknowledge that using the square of height may not have been ideal, particularly for FM and FFM, and we tried to rectify this limitation by adjusting all of the genetic associations for height.

6.5.2 Instrumental variables for body composition

This Chapter has highlighted the lack of GWASs on body composition measurements, particularly for FM and FFM (Section 6.1.3). All of the Mendelian randomization

analyses performed in this Chapter used the genetic variants that were associated with BMI at the GWS level as instruments for FMI and FFMI. Ideally, we would have used genetic variants that have been shown to be robustly associated with FM and FFM as IVs. Although there is evidence to suggest that there is significant overlap between the variants that are associated with BMI, FM and FFM, the effect that these variants have on the different body composition measurements will vary.

6.5.3 Application of multivariable methods

This Chapter has highlighted the benefits of applying multivariable methods to Mendelian randomization analyses. Despite the genetic associations between FMI and FFMI being highly correlated for both the BIA and DXA measurements, we still obtained reasonably precise estimates from the multivariable IVW method, particularly for the DXA measurements. However, the sensitivity of the multivariable methods to the strength of the correlation between the genetic associations and the correlation of the risk factors should be considered in more detail.

The discrepancy between the estimates obtained for BIA and DXA from the multivariable IVW method was unexpected. DXA is considered to be the 'gold standard' for measuring body composition, and this may have contributed to the estimates from DXA being more precise than BIA. We recommend that the analysis is re-performed when UK Biobank releases additional DXA measurements. This is particularly important as there was evidence of weak instruments under the multivariable setting for the DXA measurements.

In this Chapter, we were able to apply the multivariable MR-Egger method developed in Chapter 4 to an extensive Mendelian randomization study. As illustrated by the applied example in Chapter 4, the intercept term is sensitive to the orientation of the genetic variants. There were discrepancies in the MR-Egger intercept test for the univariable and multivariable models. For the one–sample study, the intercept was non-significant in the model for BMI, but the intercept term was significant for the DXA measurement of FMI in the multivariable MR-Egger model. This may suggest that the InSIDE assumption under the multivariable framework was not satisfied for FMI.

6.5.4 Limitations

Throughout this Chapter, we have assumed that the genetic associations from UK Biobank have been fully adjusted for ancestry by removing individuals related up to

the third degree and including the first 10 ancestry informative PCs in the regression models used to estimate the genetic associations. However, there is no guarantee that these measures would have fully accounted for population stratification. Failing to fully account for population stratification may confound the genetic association, and result in biased genetic association estimates with increased Type I error rates. Since the genetic variants selected as potential IVs were identified from an independent study to UK Biobank, the effect of having an increased Type I error rate as a result of population stratification in UK Biobank should not have biased the selection of the genetic variants. Nevertheless, if ancestry had not been fully adjusted for, then the genetic association estimates from UK Biobank may have been biased, effecting the estimates obtained from the Mendelian randomization analyses.

The genetic association estimates obtained from UK Biobank were based on the imputed genetic data, regardless of whether the genetic variant had been directly genotyped or not. Only using the imputed genetic data in the one–sample and two–sample Mendelian randomization studies could be seen as a limitation as it introduces additional uncertainty into the genetic association estimates. Furthermore, the genetic association estimates may have been overly precise as the models used to estimate the genetic associations did not account for the uncertainty in the imputation of the genetic data (Section 6.2.4). Since all of the 60 genetic variants in the liberal set had an imputation quality score > 0.96, we anticipate that the effect of using imputed genetic variants would have been minimal.

As highlighted in Section 6.2.4, the normality and homoscedasticity assumptions for the linear regression models used to estimate the genetic associations for BMI, FMI and FFMI may have been violated. This may have been rectified by transforming the data for BMI, FMI and FFMI. For ease of interpreting the estimates from the Mendelian randomization analyses, we decided not to transform BMI, FMI or FFMI. As such, the estimates and standard errors of the genetic associations with BMI, FMI and FFMI may be biased as a result of the assumptions being violated. Discrepancies in the standard errors of the genetic associations with BMI, FMI and FFMI should not affect the results from the Mendelian randomization study as they are not used in the analysis model. However, if the violation of the assumptions for linear regression effected the point estimates for the genetic associations then the Mendelian randomization estimates may have been effected as well, limiting the interpretation of the results.

6.5.5 Key points from the chapter

- In this Chapter, we investigated the effect of adiposity and body composition on asthma using univariable and multivariable Mendelian randomization analyses in the one–sample and two–sample setting.
- This is the first time that multivariable Mendelian randomization methods has been applied to this research question.
- The Mendelian randomization analyses suggested that there was a positive causal association between BMI and asthma in the one–sample setting. This was not supported by the two–sample Mendelian randomization study.
- There was some evidence from the multivariable IVW model in the one–sample study that the DXA measurement for FMI was adversely associated with asthma in adults, whereas the BIA measurement suggested that there was a null effect. However, the results were inconclusive.

Chapter 7

Conclusions and future work

The main aim of this dissertation was to develop and extend methods for Mendelian randomization analyses. Four robust methods that downweight or remove pleiotropic genetic variants were outlined in Chapter 3, MR-Egger was extended to the multivariable framework in Chapter 4, and estimating statistical interaction effects in factorial Mendelian randomization was considered in Chapter 5. These methodological developments were motivated by the applied examples within each Chapter and existing Mendelian randomization methods. Note that the methods developed in Chapters 3 and 4 both considered the issue of estimating consistent causal effects in the presence of pleiotropic genetic variants in either univariable or multivariable Mendelian randomization. Chapter 5 did not consider the implications of including pleiotropic genetic variants, but considered the possibility of estimating statistical interaction effects in Mendelian randomization.

The motivation for the work presented in Chapters 3 and 4 was also influenced by the main applied example of the dissertation: the investigation into the effect of adiposity and body composition on asthma in one–sample and two–sample Mendelian randomization studies (Chapter 6). For the one–sample and two–sample studies, the effect of BMI on asthma was considered in univariable Mendelian randomization analyses, and the effect of FMI and FFMI on asthma was considered in multivariable Mendelian randomization analyses. There was some evidence of an adverse effect of BMI on asthma, but the effect of FMI and FFMI on asthma was inconclusive. We had hoped to use the methods developed in Chapter 3 in the univariable analyses, but since the performances of the methods were not optimal, they were not included in Chapter 6. However, multivariable MR-Egger was used in Chapter 6. The factorial Mendelian randomization work was not considered, as we did not suspect that there was an interaction effect between FMI and FFMI on asthma.

We hope that this dissertation has highlighted the benefits of using Mendelian randomization to consider the causal effect of a risk factor on an outcome using observational data. Whilst the methods introduced in Chapter 3 were not optimal under all of the scenarios considered, we hope that the Multivariable MR-Egger method will be a useful addition to the literature. The work presented in Chapter 5 addresses an important gap in the Mendelian randomization literature of estimating statistical interaction effects. Finally, we hope that the application of some of the methods developed and discussed in this dissertation to the main applied example was a useful addition, despite producing inconclusive results. In the Section below, we highlight the specific contributions and limitations of the work presented in each Chapter (Section 7.1), and outline avenues of future work (Section 7.2).

7.1 Summary of the thesis

In this Section, we summarize the main findings, conclusions, and limitations of each Chapter presented in this dissertation.

7.1.1 Chapter 1

Chapter 1 provides an overview of the study designs that can detect and estimate causal effects. In particular, the Chapter introduces the concept of Mendelian randomization studies, and highlights some of the benefits and limitations of using genetic variants as IVs.

7.1.2 Chapter 2

Chapter 2 reviews the methods used in Mendelian randomization to estimate the causal effect of a risk factor on an outcome using summary level data in: 1) the primary analysis when the IV assumptions are considered to be satisfied; and 2) the sensitivity analysis when some of the genetic variants may violate the IV assumptions through pleiotropic effects. The review highlights the wide range of methods that estimate consistent causal effects when pleiotropic genetic variants are included in a Mendelian randomization analysis. The review also highlights areas that require further development: 1) identifying methods in the robust statistics literature that could be used in the sensitivity analysis of a Mendelian randomization study; and 2) developing multivariable methods that account for both measured and unmeasured

pleiotropic effects. These two limitations, and the observation of there being a lack of IV methods that estimate statistical interaction effects are considered in Chapters 3-5.

7.1.3 Chapter 3

Chapter 3 proposes four robust methods for Mendelian randomization that use summary level data. These methods are based on extensions from the robust statistics literature and robust methods in Mendelian randomization. Through an extensive simulation study and applied example, the performance of the proposed methods are compared to those frequently used in Mendelian randomization. Under certain scenarios, some of the methods proposed out perform the current methods used in the literature. Most notably, robust regression with penalized weights appears to be a worthwhile addition to sensitivity analyses when there is a small proportion of heterogeneous ratio estimates. However, in general, the weighted median estimator performed just as well, if not better, than the methods proposed in the Chapter.

The weights used in the Q-statistic were based on the NOME assumption that there was no measurement error in the genetic associations with the risk factor. However, the NOME assumption was violated in the simulation study. It is likely that the violation of the NOME assumption would have affected the performance of applying penalized weights through the Q statistic in the simulation study. The violation of the NOME assumption also limited the utility of the simulation study as it was not possible to compare the performance of the MR-Egger method with the proposed methods. Rather than using the first order weights, the modified weights proposed by Bowden *et al.* [63] that account for the violation of the NOME assumption may have been more appropriate for the simulation study.

The simulation study highlights the limitations of the four robust methods proposed in the Chapter. Most notably, there was no improvement in the performance of the methods when the number of genetic variants is increased. The methods were also sensitive to the type of heterogeneity contained in the data, with some methods working better under certain scenarios.

7.1.4 Chapter 4

Chapter 4 addresses the lack of methodology on robust methods for multivariable Mendelian randomization by extending the MR-Egger method to the multivariable setting to account for both measured and unmeasured pleiotropic effects. The assumptions required to obtain consistent causal estimates from the multivariable MR-Egger

method are considered in detail. The benefits of using multivariable MR-Egger over its univariable counterpart are also demonstrated in an extensive simulation study. Multivariable MR-Egger will be a useful addition to Mendelian randomization analyses that consider high-dimensional data where the risk factors are highly related.

As seen in the univariable version, it is not possible to determine whether the InSIDE assumption holds for multivariable MR-Egger. Furthermore, the InSIDE assumption for multivariable MR-Egger is sensitive to the correlation structure of the genetic associations and the orientation of the alleles. The recommendation of orientating the genetic variants with respect to the risk factor increasing allele is a major limitation of the method as it may result in the model being fitted multiple times in one analysis. Changing the orientation of the genetic variants will also alter the definition of the intercept term, and this may make it difficult to interpret the results if the intercept term is significant for one orientation and not another.

Since summary level data is often used in Mendelian randomization, the simulation study in Chapter 4 generated the genetic associations directly. Whilst this helped to reduce the computational burden of the simulation study, it also limited the scope of the study. By generating the summary level data directly, the impact of weak instrument bias under the multivariable setting could not be explored fully. Although the F-statistic was approximated in the simulation study, it was not possible to calculate the Sanderson-Windmeijer conditional F-statistic to consider the joint strength of the instruments.

The simulation study and findings of the Chapter were limited by the NOME assumption not being violated in the simulation study. The sensitivity of the univariable MR-Egger method to the violation of the NOME assumption has been highlighted by Bowden *et al.* [63], and is supported by the results obtained in the simulation study in Chapter 3. The impact of violating the NOME assumption, and whether the I^2 statistic is an appropriate measure to assess bias from this violation under the multivariable setting should be considered to inform applied practice.

7.1.5 Chapter 5

Chapter 5 considers the possibility of estimating statistical interaction effects in Mendelian randomization under two settings: 1) to estimate the interaction effects between two risk factors; and 2) to detect interactions between drug treatments on the risk of disease. Under the first setting, the Chapter outlines the TSLS regression model required to estimate a statistical additive interaction effect when the genetic variants are either treated as individual instruments or combined into gene scores. An extensive

simulation study shows that the power to detect the interaction term is maximised when all of the genetic variants and their interactions are included as individual IVs.

The Chapter also provides useful recommendations on using genetic variants to detect interactions between drug treatments. Although this scenario has already been considered in the applied literature, there remained unresolved methodological issues. Through simulations, the power to detect the interaction effect is considered with respect to the distribution of the gene scores, and whether the scores are treated as continuous or binary variables. The simulation study highlights the advantage of treating the gene scores as continuous variables, and the sensitivity of the method to non-symmetric distributions of the gene scores.

One of the major limitations of the work presented in the Chapter is that only two risk factors or two drug treatments can be considered in the analysis model. The applicability of the method is also limited by the risk factors and outcome only being considered with respect to continuous variables. Furthermore, the methods have only been considered with respect to statistical additive interaction effects.

We used the same criteria as multivariable Mendelian randomization to assess instrument strength. In particular, we used the Sanderson-Windmeijer conditional F-statistic to assess the joint strength of the instruments. However, it is not clear whether this is an appropriate measure for instrument strength for a factorial Mendelian randomization study.

Various factors limited the simulation studies, including: 1) the use of optimal weights in the gene scores; 2) assuming the genetic variants are uncorrelated; and 3) the lack of misspecification in the data generating model. The assumption that the genetic variants are uncorrelated particularly limits the applicability of the results presented for detecting interaction effects between drug treatments as it is likely that genetic variants in the same gene region will be in linkage disequilibrium. To consider the performance of factorial Mendelian randomization under more realistic settings, the simulations would benefit from considering additional misspecifications in the data generating model, such as misspecification of the genetic associations.

7.1.6 Chapter 6

Chapter 6 performs an extensive Mendelian randomization study to investigate the effect of adiposity and body composition on asthma using univariable and multivariable Mendelian randomization methods, including the multivariable MR-Egger method developed in Chapter 4. A one–sample Mendelian randomization study was performed using data from UK Biobank, and a two–sample Mendelian randomization study was

performed using data from UK Biobank and the GABRIEL consortium. The analyses provided some evidence of an adverse causal effect of BMI on asthma, but there was little or no evidence of a causal effect of FMI or FFMI on asthma. The DXA measurements used in UK Biobank did support an adverse effect of FMI on asthma, however, these analyses may have suffered from weak instrument bias.

The multivariable analyses may have been undermined by the genetic variants being associated with BMI, rather than being specifically associated with FMI and/or FFMI. As such, the analyses should be reperformed if results from GWASs on FMI and FFMI are released. The discrepancy in the results obtained from the analysis on adults for FMI from the BIA and DXA measurements should also be reconsidered when UK Biobank releases additional DXA measurements.

7.2 Future work

As a result of the work presented in this dissertation, we discuss potential avenues for future methodological developments in Mendelian randomization.

7.2.1 Extension to correlated genetic variants

All of the methods introduced in this dissertation assume that the genetic variants are uncorrelated. This assumption was made for two reasons: 1) for simplicity; and 2) if the validity of the IV assumptions are in doubt then it would be unwise to include correlated genetic variants. However, there will be scenarios where including correlated genetic variants would be beneficial to a Mendelian randomization study. This is particularly true for multivariable analyses where there may be few independent genetic variants that are associated with highly related risk factors. Extending the multivariable MR-Egger model to correlated genetic variants should therefore be considered. As discussed in Section 4.6, using correlated genetic variants in univariable MR-Egger should be considered first, and then extended to the multivariable setting.

The use of correlated genetic variants should also be considered for detecting interaction effects between drug treatments. As highlighted in Section 5.4.2, it is unlikely that there will be numerous independent genetic variants in the same gene region that are associated with a risk factor. It would therefore be beneficial to consider the impact of including correlated genetic variants in this setting.

7.2.2 Unresolved issues for multivariable MR-Egger

In addition to considering correlated genetic variants, there are other issues that should be addressed for multivariable MR-Egger. Perhaps one of the most pressing issues is the effect of violating the NOME assumption in the multivariable setting. Investigating the impact this violation may have would help to inform applied practice, and may be particularly important in the multivariable setting. Measurement error will bias the estimate towards the null in univariable models (as shown for MR-Egger), however, measurement error in multivariable models can bias the estimates towards or away from the null [195]. Since the violation of the NOME assumption for multivariable MR-Egger may lead to increased Type I error rates, identifying an appropriate statistic that may be able to quantify the direction and degree of bias should be considered.

One of the major limitations of the multivariable MR-Egger method is the requirement to change the orientation of the genetic variants. Bowden *et al.* [48] have rectified this issue for the univariable MR-Egger method by suggesting that radial regression be applied to the method as the estimates are independent of the orientation of the genetic variants. As such, the extension of the radial regression framework to multivariable MR-Egger should be considered.

7.2.3 Multivariable methods

The performance of multivariable methods, including the IVW method, in Mendelian randomization requires further attention. In particular, the impact of the correlation structure in the raw data and genetic associations should be investigated. The results from the multivariable models in Chapter 6 provided reasonable results despite the genetic associations being highly correlated. Perhaps the correlation structure between the risk factors, rather than the genetic associations, is more influential in the multivariable methods.

This dissertation has highlighted the lack of robust methods for multivariable Mendelian randomization. Since the application of Mendelian randomization to '-omics' data is becoming more popular, the development of robust methods that can account for measured and unmeasured pleiotropy must be considered. Although multivariable MR-Egger does contribute to this gap in the literature, additional methods are required. In particular, the literature requires methods that try to identify and remove pleiotropic genetic variants under the multivariable setting. The possibility of applying the robust methods introduced in Chapter 3, such as Lasso penalization, to the multivariable framework should be considered.

7.2.4 Interaction effects

Mendelian randomization is often compared to a RCT, with some arguing that a Mendelian randomization analysis is analogous to a 'naturally' randomized trial [196]. Although they share some similarities, the work in Chapter 5 highlighted the substantial differences in the two study designs. In particular, the difficulties in applying the compliance framework to Mendelian randomization was discussed. Investigating ways in which Mendelian randomization can be considered under a compliance framework would be a useful addition to the literature, and this may allow for the method proposed by Blackwell [137] to be applied to factorial Mendelian randomization.

Whilst Blackwell's [137] method would be a useful addition to Mendelian randomization, the suitability of the compliance framework in Mendelian randomization is limited [197, 198]. Unlike RCTs, there is no clear definition of what it means to be a complier in a Mendelian randomization study [197]. For example, if we were to perform a Mendelian randomization analysis where LDL-C was the risk factor of interest, it is not clear whether a participant would be considered as a complier to randomization if they had the LDL-C lowering variant and their LDL-C level was low, or whether we would require the LDL-C level to fall below a specified value [197]. As such, there appears to be little scope in considering Mendelian randomization under a compliance framework.

7.3 Conclusion

To conclude, the methodological developments presented in this dissertation have been motivated from applied examples. In particular, our investigation into the effect of adiposity and body composition on asthma partly motivated the work in Chapters 3 and 4. Furthermore, will have illustrated the utility of using the methods outlined in Chapters 3 to 5 in applied examples within the Chapters themselves. As such, we hope that these methods will make useful additions to the Mendelian randomization literature.

References

[1] Pearl J. *Causality*. Cambridge University Press, 2000.

[2] Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research* 2007; **16**(4):309–330, doi:10.1177/0962280206077743.

[3] Akobeng AK. Understanding randomised controlled trials. *Archives of Disease in Childhood* 2005; **90**(8):840–844, doi:10.1136/adc.2004.058222.

[4] Spieth PM, Kubasch AS, Penzlin AI, Illigens BM, Barlinn K, Siepmann T. Randomized controlled trials - a matter of design. *Neuropsychiatric Disease and Treatment* 2016; **12**:1341–1349, doi:10.2147/NDT.S101938.

[5] Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *International Journal of Epidemiology* 1992; **21**(5):837–841.

[6] Little RJ, Long Q, Lin X. A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics* 2009; **65**(2):640–649, doi:10.1111/j.1541-0420.2008.01066.x.

[7] Bennett DA, Holmes MV. Mendelian randomisation in cardiovascular research: an introduction for clinicians. *Heart* 2017; **103**(18):1400–1407, doi:10.1136/heartjnl-2016-310605.

[8] Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018; **362**:k601, doi:10.1136/bmj.k601.

[9] Jha P, Flather M, Lonn E, Farkouh M, Yusuf S. The antioxidant vitamins and cardiovascular disease. A critical review of epidemiologic and clinical trial data. *Annals of Internal Medicine* 1995; **123**(11):860–872.

[10] Khaw KT, Bingham S, Welch A, Luben R, Wareham N, Oakes S, Day N. Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: a prospective population study. European Prospective Investigation into Cancer and Nutrition. *Lancet* 2001; **357**(9257):657–663.

[11] Group HPSC. MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002; **360**(9326):23–33, doi:10.1016/S0140-6736(02)09328-5.

[12] Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 2000; **29**(4):722–729, doi:10.1093/ije/29.4.722.

[13] Sheehan NA, Didelez V. Epidemiology, genetic epidemiology and Mendelian randomisation: more need than ever to attend to detail. *Human Genetics* 2019; doi:10.1007/s00439-019-02027-3.

[14] Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* 1986; **1**(8479):507–508.

[15] Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 2003; **32**(1):1–22, doi:10.1093/ije/dyg070.

[16] Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, Davey Smith G, Sterne JA. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research* 2012; **21**(3):223–242, doi:10.1177/0962280210394459.

[17] Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; **17**(4):360–372, doi:10.1097/01.ede.0000222409.00878.37.

[18] Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *International Journal of Epidemiology* 2013; **42**(5):1497–1501, doi:10.1093/ije/dyt179.

[19] Freeman G, Cowling BJ, Schooling CM. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *International Journal of Epidemiology* 2013; **42**(4):1157–1163, doi:10.1093/ije/dyt110.

[20] Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology* 2004; **33**(1):30–42, doi:10.1093/ije/dyh132.

[21] Sekula P, Del Greco M F, Pattaro C, Kottgen A. Mendelian Randomization as an Approach to Assess Causality Using Observational Data. *Journal of the American Society of Nephrology* 2016; **27**(11):3253–3265, doi:10.1681/ASN.2016010098.

[22] Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics* 2018; **27**(R2):R195–R208, doi:10.1093/hmg/ddy163.

[23] Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, Evans DM, Davey Smith G. Recent Developments in Mendelian Randomization Studies. *Current Epidemiology Reports* 2017; **4**(4):330–345, doi:10.1007/s40471-017-0128-6.

[24] Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. *International Journal of Epidemiology* 2016; **45**(3):908–915, doi:10.1093/ije/dyw127.

[25] Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *International Journal of Epidemiology* 2016; **45**(6):1717–1726, doi:10.1093/ije/dyx028.

[26] Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, Paul DS, Freitag D, Burgess S, Danesh J, *et al.*. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 2016; **32**(20):3207–3209, doi:10.1093/bioinformatics/btw373.

[27] Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, *et al.*. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018; **7**:e34 408, doi:10.7554/eLife.34408.

[28] Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *American Journal of Epidemiology* 2015; **181**(4):251–260, doi:10.1093/aje/kwu283.

[29] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* 2015; **44**(2):512–525, doi:10.1093/ije/dyv080.

[30] Burgess S, Bowden J, Dudbridge F, Thompson S. Robust instrumental variable methods using multiple candidate instruments with application to Mendelian randomization. *ArXiv e-prints* 2016; .

[31] Rees JMB, Wood AM, Dudbridge F, Burgess S. Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates. *PLoS ONE* 2019; **14**(9):e0222 362, doi:10.1371/journal.pone.0222362.

[32] Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Statistics in Medicine* 2017; **36**(29):4705–4718, doi:10.1002/sim.7492.

[33] Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2 × 2 factorial Mendelian randomization study. *Journal of the American College of Cardiology* 2015; **65**(15):1552–1561, doi:10.1016/j.jacc.2015.02.020.

[34] Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, Voros S, Giugliano RP, Davey Smith G, Fazio S, *et al.*. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. *New England Journal of Medicine* 2016; **375**(22):2144–2153, doi:10.1056/NEJMoa1604304.

[35] Ference BA, Kastelein JJP, Ginsberg HN, Chapman MJ, Nicholls SJ, Ray KK, Packard CJ, Laufs U, Brook RD, Oliver-Williams C, *et al.*. Association of Genetic Variants Related to CETP Inhibitors and Statins With Lipoprotein Levels and Cardiovascular Risk. *Journal of the American Medical Association* 2017; **318**(10):947–956, doi:10.1001/jama.2017.11467.

[36] Rees JMB, Foley CN, Burgess S. Factorial Mendelian randomization: using genetic variants to assess interactions. *International Journal of Epidemiology* 2019; doi:10.1093/ije/dyz161.

[37] Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E, Farrall M, Lathrop M, Cookson WOCM, *et al.*. A large-scale, consortium-based genomewide association study of asthma. *New England Journal of Medicine* 2010; **363**(13):1211–1221, doi:10.1056/NEJMoa0906312.

[38] Angrist JD, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 1995; **90**(430):431–442, doi:10.1080/01621459.1995.10476535.

[39] Anderson TW, Rubin H. Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics* 1949; **20**(1):46–63, doi:10.1214/aoms/1177730090.

[40] Baum C, Schaffer M, Stillman S. Instrumental variables and gmm: Estimation and testing. *Stata Journal* 2003; **3**(1):1–31.

[41] Johnston KM, Gustafson P, Levy AR, Grootendorst P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Statistics in Medicine* 2008; **27**(9):1539–1556, doi:10.1002/sim.3036.

[42] Frank K, Eric Z. Bayesian and classical approaches to instrumental variable regression. *Journal of Econometrics* 2003; **114**(1):29–72, doi:10.1016/S0304-4076(02)00219-1.

[43] Berzuini C, Guo H, Burgess S, Bernardinelli L. A Bayesian approach to Mendelian randomization with multiple pleiotropic variants. *Biostatistics* 2018; :kxy027doi:10.1093/biostatistics/kxy027.

[44] Johnson T. Efficient calculation for multi-SNP genetic risk scores. *Technical Report*, The Comprehensive R Archive Network 2013. Available at http://cran.r-project.org/web/packages/gtx/vignettes/ashg2012.pdf [last accessed 2016/06/14].

[45] Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology* 2013; **37**(7):658–665, doi:10.1002/gepi.21758.

[46] Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *European Journal of Epidemiology* 2017; **32**(5):377–389, doi:10.1007/s10654-017-0255-x.

[47] Thomas DC, Lawlor DA, Thompson JR. Re: Estimation of bias in nongenetic observational studies using "Mendelian triangulation" by Bautista et al. *Annals Epidemiology* 2007; **17**(7):511–513, doi:10.1016/j.annepidem.2006.12.005.

[48] Bowden J, Spiller W, Del Greco M F, Sheehan N, Thompson J, Minelli C, Davey Smith G. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *International Journal of Epidemiology* 2018; **47**(4):1264–1278, doi:10.1093/ije/dyy101.

[49] Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I^2 statistic. *International Journal of Epidemiology* 2016; **45**(6):1961–1974, doi:10.1093/ije/dyw220.

[50] Thompson S, Sharp S. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999; **18**(20):2693–2708.

[51] Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Statistics in Medicine* 2017; **36**(11):1783–1802, doi:10.1002/sim.7221.

[52] Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic Epidemiology* 2016; **40**(4):304–314, doi:10.1002/gepi.21965.

[53] Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Statistical Science* 1999; **14**(1):29–46, doi:10.1214/ss/1009211805.

[54] Palmer TM, Sterne JA, Harbord RM, Lawlor DA, Sheehan NA, Meng S, Granell R, Davey Smith G, Didelez V. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. *American Journal of Epidemiology* 2011; **173**(12):1392–1403, doi:10.1093/aje/kwr026.

[55] Didelez V, Meng S, Sheehan NA. Assumptions of iv methods for observational epidemiology. *Statistical Science* 2010; **25**(1):22–40, doi:10.1214/09-STS316.

[56] Vansteelandt S, Bowden J, Babanezhad M, Goetghebeur E. On instrumental variables estimation of causal odds ratios. *Statistical Science* 2011; **26**(3):403–422.

[57] von Hinke S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. Genetic markers as instrumental variables. *Journal of Health Economics* 2016; **45**:131–148, doi:PMC4770870.

[58] Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, Carpenter J, Rucker G, Harbord RM, Schmid CH, *et al.*. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011; **343**:d4002, doi:10.1136/bmj.d4002.

[59] Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. *Epidemiology* 2017; **28**(1):30–42, doi:10.1097/EDE.0000000000000559.

[60] Galbraith RF. A note on graphical presentation of estimated odds ratios from several clinical trials. *Statistics in Medicine* 1988; **7**(8):889–894.

[61] Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954; **10**:101–29.

[62] Del Greco M F, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Statistics in Medicine* 2015; **34**(21):2926–2940, doi:10.1002/sim.6522.

[63] Bowden J, Del Greco M F, Minelli C, Lawlor D, Sheehan N, Thompson J, Davey Smith G. Improving the accuracy of two-sample summary data mendelian randomization: moving beyond the NOME assumption. *bioRxiv* 2017; doi:10.1101/159442.

[64] Kolesár M, Chetty R, Friedman J, Glaeser E, Imbens G. Identification and inference with many invalid instruments. *Journal of Business & Economic Statistics* 2015; **33**(4):474–484, doi:10.1080/07350015.2014.978175.

[65] Kang H, Zhang A, Cai T, Small D. Instrumental Variables Estimation With Some Invalid Instruments and its Application to Mendelian Randomization. *Journal of the American Statistical Association* 2016; **111**(513):132–144, doi:10.1080/01621459.2014.994705.

[66] Guo Z, Kang H, Tony Cai T, Small DS. Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2018; **80**(4):793–815, doi:10.1111/rssb.12275.

[67] Windmeijer F, Farbmacher H, Davies N, Davey Smith G. On the Use of the Lasso for Instrumental Variables Estimation with Some Invalid Instruments. *Journal of the American Statistical Association* 2018; **0**(0):1–12, doi:10.1080/01621459.2018.1498346.

[68] van Kippersluis H, Rietveld CA. Pleiotropy-robust Mendelian randomization. *International Journal of Epidemiology* 2018; **47**(4):1279–1288, doi:PMC6124631.

[69] Spiller W, Slichter D, Bowden J, Davey Smith G. Detecting and correcting for bias in Mendelian randomization analyses using Gene-by-Environment interactions. *International Journal of Epidemiology* 2018; doi:10.1093/ije/dyy204.

[70] Detecting invalid instruments using L1-GMM. *Economics Letters* 2008; **101**(3):285 – 287, doi:https://doi.org/10.1016/j.econlet.2008.09.004.

[71] Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology* 2017; **46**(6):1985–1998, doi:10.1093/ije/dyx102.

[72] Burgess S, Zuber V, Gkatzionis A, Foley CN. Modal-based estimation via heterogeneity-penalized weighting: model averaging for consistent and efficient

estimation in Mendelian randomization when a plurality of candidate instruments are valid. *International Journal of Epidemiology* 2018; **47**(4):1242–1254, doi:10.1093/ije/dyy080.

[73] Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics* 2018; **50**(5):693–698, doi:10.1038/s41588-018-0099-7.

[74] Dai JY, Peters U, Wang X, Kocarnik J, Chang-Claude J, Slattery ML, Chan A, Lemire M, Berndt SI, Casey G, *et al.*. Diagnostics for Pleiotropy in Mendelian Randomization Studies: Global and Individual Tests for Direct Effects. *American Journal of Epidemiology* 2018; **187**(12):2672–2680, doi:10.1093/aje/kwy177.

[75] Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, Robinson MR, McGrath JJ, Visscher PM, Wray NR, *et al.*. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications* 2018; **9**(1):224, doi:10.1038/s41467-017-02317-2.

[76] Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**(7109):629–634, doi:PMC2127453.

[77] Sawa T. The exact sampling distribution of ordinary least squares and two-stage least squares estimators. *Journal of the American Statistical Association* 1969; **64**(327):923–937.

[78] Nelson C, Startz R. The Distribution of the Instrumental Variables Estimator and its t-Ratio When the Instrument is a Poor One. *The Journal of Business* 1990; **63**(1):S125–S140.

[79] Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Statistics in Medicine* 2011; **30**(11):1312–1323, doi:10.1002/sim.4197.

[80] Burgess S, Dudbridge F, Thompson SG. Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". *American Journal of Epidemiology* 2015; **181**(4):290–291, doi:10.1093/aje/kwv017.

[81] Friedewald W, Levy R, Fredrickson D. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clinical Chemistry* 1972; **18**(6):499–502.

[82] Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable mendelian randomization in the single sample and two-sample summary data settings. *International Journal of Epidemiology* 2018; .

[83] Sanderson E, Windmeijer F. A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics* 2016; **190**(2):212 – 221, doi:10.1016/j.jeconom.2015.06.004.

[84] Huber P. *Robust Statistics*. Wiley, 2009.

[85] Cheng X, Liao Z. Select the valid and relevant moments: An information-based LASSO for GMM with many moments. *Journal of Econometrics* 2015; **186**(2):443 – 464, doi:10.1016/j.jeconom.2015.02.019.

[86] RStudio Team. *RStudio: Integrated Development Environment for R*.

[87] Hampel FR. A general qualitative definition of robustness. *The Annals of Mathematical Statistics* 1971; **42**(6):1887–1896.

[88] Koller M, Stahel W. Sharpening wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis* 2011; **55**(8):2504–2515, doi:10.1016/j.csda.2011.02.014.

[89] Yu C, Yao W. Robust linear regression: A review and comparison. *Communications in Statistics - Simulation and Computation* 2017; **46**(8):6261–6282, doi:10.1080/03610918.2016.1202271.

[90] Almetwally E, Almongy H. Comparison between m-estimation, s-estimation, and mm estimation methods of robust estimation with application and simulation. *International Journal of Mathematical Archive* 2018; **11**(9):55–63.

[91] Salibian-Barrera M, Yohai VJ. A fast algorithm for s-regression estimates. *Journal of Computational and Graphical Statistics* 2006; **15**(2):414–427, doi:10.1198/106186006X113629.

[92] Yohai VJ. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics* 1987; **15**(2):642–656, doi:10.1214/aos/1176350366.

[93] Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, Koller M, Maechler M. robustbase: Basic Robust Statistics 2015; URL http://cran.r-project.org/package=robustbase, R package version 0.92-5.

[94] Bowden J, Hemani G, Davey Smith G. Invited Commentary: Detecting Individual and Global Horizontal Pleiotropy in Mendelian Randomization—A Job for the Humble Heterogeneity Statistic? *American Journal of Epidemiology* 2018; **187**(12):2681–2685, doi:10.1093/aje/kwy185.

[95] Rucker G, Schwarzer G, Carpenter JR, Binder H, Schumacher M. Treatment effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics* 2011; **12**(1):122–142, doi:10.1093/biostatistics/kxq046.

[96] Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society, Series B (Methodological)* 1996; **58**:267–288.

[97] Belloni A, Chen DL, Chernozhukov V, Hansen C. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 2012; **80**(6):2369–2429.

[98] Goeman J, Meijer R, Chaturvedi N, Lueder M. penalized: L1 (Lasso and Fused Lasso) and L2 (Ridge) Penalized Estimation in GLMs and in the Cox Model 2017; URL https://cran.r-project.org/web/packages/penalized/penalized.pdf.

[99] Rousseeuw PJ. Least median of squares regression. *Journal of the American Statistical Association* 1984; **79**:871–880, doi:10.1080/01621459.1984.10477105.

[100] Rousseeuw PJ, Leroy AM. *Robust regression and outlier detection*. Wiley, 1987.

[101] Pison G, Van Aelst S, Willems G. Small sample corrections for LTS and MCD. *Metrika* 2002; **55**:111–123, doi:10.1007/s001840200191.

[102] Coodin S. Body mass index in persons with schizophrenia. *Canadian Journal of Psychiatry* 2001; **46**(6):549–555, doi:10.1177/070674370104600610.

[103] Allison DB, Fontaine KR, Heo M, Mentore JL, Cappelleri JC, Chandler LP, Weiden PJ, Cheskin LJ. The distribution of body mass index among individuals with and without schizophrenia. *Journal of Clinical Psychiatry* 1999; **60**(4):215–220.

[104] Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, *et al.*. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**(7538):197–206, doi:10.1038/nature14177.

[105] Ripke S, Neale BM, Corvin A, Walters JT, Farh KH, *et al.*. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**(7510):421–427, doi:10.1038/nature13595.

[106] Hartwig FP, Bowden J, Loret de Mola C, Tovo-Rodrigues L, Davey Smith G, Horta BL. Body mass index and psychiatric disorders: a Mendelian randomization study. *Scientific Reports* 2016; **6**:32 730, doi:10.1038/srep32730.

[107] Notkola IL, Sulkava R, Pekkanen J, Erkinjuntti T, Ehnholm C, Kivinen P, Tuomilehto J, Nissinen A. Serum total cholesterol, apolipoprotein E epsilon 4 allele, and Alzheimer's disease. *Neuroepidemiology* 1998; **17**(1):14–20.

[108] Solomon A, Kareholt I, Ngandu T, Winblad B, Nissinen A, Tuomilehto J, Soininen H, Kivipelto M. Serum cholesterol changes after midlife and late-life cognition: twenty-one-year follow-up study. *Neurology* 2007; **68**(10):751–756, doi:10.1212/01.wnl.0000256368.57375.b7.

[109] Shepardson NE, Shankar GM, Selkoe DJ. Cholesterol level and statin use in Alzheimer disease: I. Review of epidemiological and preclinical studies. *JAMA Neurology* 2011; **68**(10):1239–1244, doi:10.1001/archneurol.2011.203.

[110] Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, *et al.*. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nature Genetics* 2013; **45**(11):1345–1352, doi:10.1038/ng.2795.

[111] Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, *et al.*. Discovery and refinement of loci associated with lipid levels. *Nature Genetics.* 2013; **45**(11):1274–1283, doi:10.1038/ng.2797.

[112] Benn M, Nordestgaard BG, Frikke-Schmidt R, Tybjærg-Hansen A. Low LDL cholesterol, PCSK9 and HMGCR genetic variation, and risk of Alzheimer's disease and Parkinson's disease: Mendelian randomisation study. *British Medical Journal* 2017; **357**, doi:10.1136/bmj.j1648.

[113] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, *et al.*. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* 2013; **45**(12):1452–1458, doi:10.1038/ng.2802.

[114] Helgadottir A, Gretarsdottir S, Thorleifsson G, Hjartarson E, Sigurdsson A, *et al.*. Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. *Nature Genetics* 2016; **48**(6):634–639, doi:10.1038/ng.3561.

[115] Burgess S, Dudbridge F, Thompson SG. Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". *American Journal of Epidemiology* 2015; **181**(4):290–291, doi:10.1093/aje/kwv017.

[116] Burgess S, Harshfield E. Mendelian randomization to assess causal effects of blood lipids on coronary heart disease: lessons from the past and applications to the future. *Current Opinion Endocrinology, Diabetes and Obesity* 2016; **23**(2):124–130, doi:10.1097/MED.0000000000000230.

[117] Burgess S, Freitag DF, Khan H, Gorman DN, Thompson SG. Using multivariable Mendelian randomization to disentangle the causal effects of lipid fractions. *PLOS One* 2014; **9**(10):e108 891, doi:10.1371/journal.pone.0108891.

[118] Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, *et al.*. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics* 2015; **47**(10):1121–1130, doi:10.1038/ng.3396.

[119] Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, Ray K, Thompson A, Wood A, Lewington S, Sattar N. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* 2009; **302**(18):1993–2000, doi:10.1001/jama.2009.1619.

[120] White J, Sofat R, Hemani G, Shah T, Engmann J, *et al.*. Plasma urate concentration and risk of coronary heart disease: a Mendelian randomisation analysis. *The Lancet Diabetes and Endocrinology* 2016; **4**(4):327–336, doi:10.1016/S2213-8587(15)00386-1.

[121] Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Statistics in Medicine* 2016; **35**(11):1880–1906, doi:10.1002/sim.6835.

[122] Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors. *Journal of Clinical Epidemiology* 2015; **69**:208–216, doi:10.1016/j.jclinepi.2015.08.001.

[123] Fritz M, Mackinnon D. Required sample size to detect the mediated effect. *Psychological Science* 2007; **18**(3):233–239, doi:10.1111/j.1467-9280.2007.01882.x.

[124] Relton CL, Davey Smith G. Is epidemiology ready for epigenetics? *International Journal of Epidemiology* 2012; **41**(1):5–9, doi:10.1093/ije/dys006.

[125] Suhre K, Raffler J, Kastenmuller G. Biochemical insights from population studies with genetics and metabolomics. *Archives of Biochemistry and Biophysics* 2016; **589**:168–176, doi:10.1016/j.abb.2015.09.023.

[126] Kastenmuller G, Raffler J, Gieger C, Suhre K. Genetics of human metabolism: an update. *Human Molecular Genetics* 2015; **24**(R1):R93–R101, doi:10.1093/hmg/ddv263.

[127] Montgomery AA, Peters TJ, Little P. Design, analysis and presentation of factorial randomised controlled trials. *BMC Medical Research Methodology* 2003; **3**:26, doi:10.1186/1745-6215-12-179.

[128] VanderWeele TJ, Knol MJ. A Tutorial on Interaction. *Technical Report* 2014. Available at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/603/2013/03/InteractionTutorial.pdf [last accessed 2017/10/26].

[129] McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. *Journal of the American Medical Association* 2003; **289**(19):2545–2553, doi:10.1001/jama.289.19.2545.

[130] Dakin H, Gray A. Economic evaluation of factorial randomised controlled trials: challenges, methods and recommendations. *Statistics in Medicine* 2017; **36**(18):2814–2830, doi:10.1002/sim.7322.

[131] Iacobucci D, Schneider MJ, Popovich DL, Bakamitsos GA. Mean centering helps alleviate "micro" but not "macro" multicollinearity. *Behavior Research Methods* 2016; **48**(4):1308–1317, doi:10.3758/s13428-015-0624-x.

[132] Cragg JG, Donald SG. Testing identifiability and specification in instrumental variable models. *Econometric Theory* 1993; **9**(2):222–240, doi:10.1017/S0266466600007519.

[133] Wooldridge JM. *Econometric analysis of cross section and panel data*. MIT Press, 2002.

[134] VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology* 2009; **20**(6):863–871, doi:10.1097/EDE.0b013e3181ba333c.

[135] Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics* 2014; **23**(R1):89–98, doi:10.1093/hmg/ddu328.

[136] Carter AR, Borges M, Benn M, Tybjærg-Hansen A, Davey Smith G, Nordestgaard BG, Lawlor DA. Combined Association of Body Mass Index and Alcohol Consumption With Biomarkers for Liver Injury and Incidence of Liver Disease: A Mendelian Randomization Study. *JAMA Network Open* 2019; **2**(3):e190 305–e190 305, doi:10.1001/jamanetworkopen.2019.0305.

[137] Blackwell M. Instrumental variable methods for conditional effects and causal interaction in voter mobilization experiments. *Journal of the American Statistical Association* 2017; **112**(518):590–599, doi:10.1080/01621459.2016.1246363.

[138] Brittain E, Wittes J. Factorial designs in clinical trials: The effects of non-compliance and subadditivity. *Statistics in Medicine* 1989; **8**(2):161–171, doi:10.1002/sim.4780080204.

[139] Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Statistical Methods in Medical Research* 2017; **26**(5):2333–2355, doi:10.1177/0962280215597579.

[140] Jiang Y, Small D. ivpack: Instrumental Variable Estimation 2014; URL https://CRAN.R-project.org/package=ivpack.

[141] Wickham H. ggplot2: Elegant Graphics for Data Analysis 2016; URL https://ggplot2.tidyverse.org.

[142] Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *European Journal of Epidemiology* 2015; **30**(7):543–552, doi:10.1007/s10654-015-0011-z.

[143] Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002; **360**(9349):1903–1913.

[144] Wei GS, Coady SA, Goff DC, Brancati FL, Levy D, Selvin E, Vasan RS, Fox CS. Blood pressure and the risk of developing diabetes in african americans and whites: ARIC, CARDIA, and the framingham heart study. *Diabetes Care* 2011; **34**(4):873–879, doi:10.2337/dc10-1786.

[145] Droyvold WB, Midthjell K, Nilsen TI, Holmen J. Change in body mass index and its impact on blood pressure: a prospective population study. *International Journal of Obesity* 2005; **29**(6):650–655, doi:10.1038/sj.ijo.0802944.

[146] Gelber RP, Gaziano JM, Manson JE, Buring JE, Sesso HD. A prospective study of body mass index and the risk of developing hypertension in men. *American Journal of Hypertension* 2007; **20**(4):370–377, doi:10.1016/j.amjhyper.2006.10.011.

[147] Roerecke M, Kaczorowski J, Tobe SW, Gmel G, Hasan OSM, Rehm J. The effect of a reduction in alcohol consumption on blood pressure: a systematic review and meta-analysis. *Lancet Public Health* 2017; **2**(2):e108–e120, doi:10.1016/S2468-2667(17)30003-8.

[148] Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, Ntalla I, Surendran P, Liu C, Cook JP, *et al.*. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nature Genetics* 2017; **49**(3):403–415, doi:10.1038/ng.3768.

[149] Lewis SJ, Zuccolo L, Davey Smith G, Macleod J, Rodriguez S, Draper ES, Barrow M, Alati R, Sayal K, Ring S, *et al.*. Fetal alcohol exposure and IQ at age 8: evidence from a population-based birth-cohort study. *PLoS ONE* 2012; **7**(11):e49 407, doi:10.1371/journal.pone.0049407.

[150] Clifton EA, Day FR, De Lucia Rolfe E, Forouhi NG, Brage S, Griffin SJ, Wareham NJ, Ong KK. Associations between body mass index-related genetic variants and adult body composition: The Fenland cohort study. *International Journal of Obesity* 2017; **41**(4):613–619, doi:10.1038/ijo.2017.11.

[151] Simpson CR, Sheikh A. Trends in the epidemiology of asthma in England: a national study of 333,294 patients. *Journal of the Royal Society of Medicine* 2010; **103**(3):98–106, doi:10.1258/jrsm.2009.090348.

[152] Anderson HR, Gupta R, Strachan DP, Limb ES. 50 years of asthma: UK trends from 1955 to 2004. *Thorax* 2007; **62**(1):85–90, doi:10.1136/thx.2006.066407.

[153] Mukherjee M, Stoddart A, Gupta RP, Nwaru BI, Farr A, Heaven M, Fitzsimmons D, Bandyopadhyay A, Aftab C, Simpson CR, *et al.*. The epidemiology, healthcare and societal burden and costs of asthma in the UK and its member nations: analyses of standalone and linked national databases. *BMC Medicine* 2016; **14**(1):113, doi:10.1186/s12916-016-0657-8.

[154] Eder W, Ege MJ, von Mutius E. The asthma epidemic. *New England Journal of Medicine* 2006; **355**(21):2226–2235, doi:10.1056/NEJMra054308.

[155] UK A. Asthma facts and statistics. *Technical Report* 2018. Https://www.asthma.org.uk/about/media/facts-and-statistics/ [last accessed 2018/11/1].

[156] Figueroa-Munoz JI, Chinn S, Rona RJ. Association between obesity and asthma in 4-11 year old children in the UK. *Thorax* 2001; **56**(2):133–137, doi:10.1136/thorax.56.2.133.

[157] Wickens K, Barry D, Friezema A, Rhodius R, Bone N, Purdie G, Crane J. Obesity and asthma in 11-12 year old New Zealand children in 1989 and 2000. *Thorax* 2005; **60**(1):7–12, doi:10.1136/thx.2002.001529.

[158] Rzehak P, Wijga AH, Keil T, Eller E, Bindslev-Jensen C, Smit HA, Weyler J, Dom S, Sunyer J, Mendez M, *et al.*. Body mass index trajectory classes and incident asthma in childhood: results from 8 European Birth Cohorts–a Global Allergy and Asthma European Network initiative. *Journal of Allergy and Clinical Immunology* 2013; **131**(6):1528–1536, doi:10.1016/j.jaci.2013.01.001.

[159] Gilliland FD, Berhane K, Islam T, McConnell R, Gauderman WJ, Gilliland SS, Avol E, Peters JM. Obesity and the risk of newly diagnosed asthma in school-age children. *American Journal of Epidemiology* 2003; **158**(5):406–415.

[160] Wang R, Custovic A, Simpson A, Belgrave DC, Lowe LA, Murray CS. Differing associations of BMI and body fat with asthma and lung function in children. *Pediatric Pulmonology* 2014; **49**(11):1049–1057, doi:10.1002/ppul.22927.

[161] Luder E, Ehrlich RI, Lou WY, Melnik TA, Kattan M. Body mass index and the risk of asthma in adults. *Respiratory Medicine* 2004; **98**(1):29–37, doi:10.1016/j.rmed.2003.08.004.

[162] Fenger RV, Gonzalez-Quintela A, Vidal C, Husemoen LL, Skaaby T, Thuesen BH, Aadahl M, Madsen F, Linneberg A. The longitudinal relationship of changes of adiposity to changes in pulmonary function and risk of asthma in a general adult population. *BMC Pulmonary Medicine* 2014; **14**:208, doi:10.1186/1471-2466-14-208.

[163] Beuther DA, Sutherland ER. Overweight, obesity, and incident asthma: a meta-analysis of prospective epidemiologic studies. *American Journal of Respiratory and Critical Care Medicine* 2007; **175**(7):661–666, doi:10.1164/rccm.200611-1717OC.

[164] Granell R, Henderson AJ, Evans DM, Davey Smith G, Ness AR, Lewis S, Palmer TM, Sterne JA. Effects of BMI, fat mass, and lean mass on asthma in childhood: a Mendelian randomization study. *PLoS Medicine* 2014; **11**(7):e1001 669, doi:10.1371/journal.pmed.1001669.

[165] Skaaby T, Taylor AE, Thuesen BH, Jacobsen RK, Friedrich N, Mollehave LT, Hansen S, Larsen SC, Volker U, Nauck M, *et al.*. Estimating the causal effect of body mass index on hay fever, asthma and lung function using Mendelian randomization. *Allergy* 2017; **73**(1):153–164, doi:10.1111/all.13242.

[166] Daniels SR, Khoury PR, Morrison JA. The utility of body mass index as a measure of body fatness in children and adolescents: differences by race and gender. *Pediatrics* 1997; **99**(6):804–807.

[167] Willett WC, Dietz WH, Colditz GA. Guidelines for healthy weight. *New England Journal of Medicine* 1999; **341**(6):427–434, doi:10.1056/NEJM199908053410607.

[168] Romero-Corral A, Lopez-Jimenez F, Sierra-Johnson J, Somers VK. Differentiating between body fat and lean mass-how should we measure obesity? *Nature Reviews Endocrinology* 2008; **4**(6):322–323, doi:10.1038/ncpendmet0809.

[169] Yu S, Visvanathan T, Field J, Ward LC, Chapman I, Adams R, Wittert G, Visvanathan R. Lean body mass: the development and validation of prediction equations in healthy adults. *BMC Pharmacology and Toxicology* 2013; **14**:53.

[170] Scafoglieri A, Clarys JP. Dual energy X-ray absorptiometry: gold standard for muscle mass? *J Cachexia Sarcopenia Muscle* 2018; **9**(4):786–787.

[171] Andreoli A, Scalzo G, Masala S, Tarantino U, Guglielmi G. Body composition assessment by dual-energy X-ray absorptiometry (DXA). *Musculoskeletal Radiology* 2009; **114**(2):286–300, doi:10.1007/s11547-009-0369-7.

[172] Achamrah N, Colange G, Delay J, Rimbert A, Folope V, Petit A, Grigioni S, Dechelotte P, Coeffier M. Comparison of body composition assessment by DXA and BIA according to the body mass index: A retrospective study on 3655 measures. *PLOS ONE* 2018; **13**(7):1–13, doi:10.1371/journal.pone.0200465.

[173] Sillanpaa E, Cheng S, Hakkinen K, Finni T, Walker S, Pesola A, Ahtiainen J, Stenroth L, Selanne H, Sipila S. Body composition in 18- to 88-year-old adults–comparison of multifrequency bioimpedance and dual-energy X-ray absorptiometry. *Obesity* 2014; **22**(1):101–109, doi:10.1002/oby.20583.

[174] Lee SY, Gallagher D. Assessment methods in human body composition. *Current Opinion in Clinical Nutrition and Metabolic Care* 2008; **11**(5):566–572, doi:10.1097/MCO.0b013e32830b5f23.

[175] Sandholt CH, Grarup N, Pedersen O, Hansen T. Genome-wide association studies of human adiposity: Zooming in on synapses. *Molecular and Cellular Endocrinology* 2015; **418 Pt 2**:90–100, doi:10.1016/j.mce.2015.09.029.

[176] Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, *et al.*. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007; **316**(5826):889–894, doi:10.1126/science.1141634.

[177] Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J, Najjar S, Nagaraja R, Orru M, Usala G, *et al.*. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genetics* 2007; **3**(7):e115, doi:10.1371/journal.pgen.0030115.

[178] Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Lango Allen H, Lindgren CM, Luan J, Magi R, *et al.*. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics* 2010; **42**(11):937–948, doi:10.1038/ng.686.

[179] Pei YF, Ren HG, Liu L, Li X, Fang C, Huang Y, Hu WZ, Kong WW, Feng AP, You XY, *et al.*. Genomic variants at 20p11 associated with body fat mass in the European population. *Obesity* 2017; **25**(4):757–764, doi:10.1002/oby.21775.

[180] Zillikens MC, Demissie S, Hsu YH, Yerges-Armstrong LM, Chou WC, Stolk L, Livshits G, Broer L, Johnson T, Koller DL, *et al.*. Large meta-analysis of genome-wide association studies identifies five loci for lean body mass. *Nature Communications* 2017; **8**(1):80, doi:10.1038/s41467-017-00031-7.

[181] March ME, Sleiman PM, Hakonarson H. Genetic polymorphisms and associated susceptibility to asthma. *International Journal of General Medicine* 2013; **6**:253–265, doi:10.2147/IJGM.S28156.

[182] Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genetic Epidemiology* 2016; **40**(7):597–608, doi:10.1002/gepi.21998.

[183] Hartwig FP, Davies NM. Why internal weights should be avoided (not only) in MR-Egger regression. *International Journal of Epidemiology* 10 2016; **45**(5):1676–1678, doi:10.1093/ije/dyw240.

[184] Palmer LJ. UK Biobank: bank on it. *Lancet* Jun 2007; **369**(9578):1980–1982, doi:10.1016/S0140-6736(07)60924-6.

[185] Thompson SG, Willeit P. UK Biobank comes of age. *Lancet* 2015; **386**(9993):509–510, doi:10.1016/S0140-6736(15)60578-5.

[186] Biobank U. Uk biobank touchscreen questionnaire. *Technical Report* 2007. Available at https://www.ukbiobank.ac.uk/wp-content/uploads/2011/06/Touch_screen_questionnaire.pdf [last accessed 2018/10/30].

[187] Biobank U. Verbal interview stage. *Technical Report* 2012. Https://biobank.ctsu.ox.ac.uk/crystal/docs/Interview.pdf [last accessed 2018/10/30].

[188] Biobank U. Protocol for a large-scale prospective epidemiological resource. *Technical Report* 2007. Available at http://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf [last accessed 2018/10/30].

[189] Harvey NC, Matthews P, Collins R, Cooper C, Cooper C, Harvey N, Compston J, Eastell R, Francis RM, Ralston SH, *et al.*. Osteoporosis epidemiology in UK Biobank: a unique opportunity for international researchers. *Osteoporosis International* 2013; **24**(12):2903–2905, doi:10.1007/s00198-013-2508-1.

[190] Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA, *et al.*. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* 2016; **167**(5):1415–1429.

[191] Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, Westra HJ, Shakhbazov K, Abdellaoui A, Agrawal A, *et al.*. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 2013; **340**(6139):1467–1471, doi:10.1126/science.1235488.

[192] Furberg H, Kim Y, Dackor J, Boerwinkle E, Franceschini N, Ardissino D, Bernardinelli L, Mannucci PL, Mauri F, Merlini PA, *et al.*. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics* 2010; **42**(5):441–447, doi:10.1038/ng.571.

[193] Duffy GWD, Qiu MMW, Lazarus R. *GeneticsDesign: Functions for designing genetics studies* 2019. R package version 1.52.0.

[194] Stergiakouli E, Gaillard R, Tavare JM, Balthasar N, Loos RJ, Taal HR, Evans DM, Rivadeneira F, St Pourcain B, Uitterlinden AG, *et al.*. Genome-wide association study of height-adjusted BMI in childhood identifies functional variant in ADCY3. *Obesity* 2014; **22**(10):2252–2259, doi:10.1002/oby.20840.

[195] Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American Journal of Epidemiology* 1990; **132**(4):734–745.

[196] Thanassoulis G, O'Donnell CJ. Mendelian randomization: nature's randomized trial in the post-genome era. *JAMA* 2009; **301**(22):2386–2388, doi:10.1001/jama.2009.812.

[197] Swanson SA, Tiemeier H, Ikram MA, Hernan MA. Nature as a Trialist?: Deconstructing the Analogy Between Mendelian Randomization and Randomized Trials. *Epidemiology* 2017; **28**(5):653–659, doi:10.1097/EDE.0000000000000699.

[198] Swanson SA, Hernan MA. The challenging interpretation of instrumental variable estimates under monotonicity. *Internatioanl Journal of Epidemiology* 2018; **47**(4):1289–1297, doi:10.1093/ije/dyx038.

Appendix A

Paper from arXiv: Robust instrumental variable methods using multiple candidate instruments with application to Mendelian randomization

Preliminary paper on the work that was adapted and extended on in Chapter 3 that was uploaded to arXiv by Burgess *et al.* [30] on 12th June 2016.

Assessing the effectiveness of robust instrumental variable methods using multiple candidate instruments with application to Mendelian randomization

Stephen Burgess [1, *] Jack Bowden [2] Frank Dudbridge [1,3]
Simon G. Thompson [1]

[1] Department of Public Health and Primary Care,
University of Cambridge, Cambridge, UK.

[2] Integrative Epidemiology Unit,
University of Bristol, Bristol, UK.

[3] Department of Non-communicable Disease Epidemiology,
London School of Hygiene and Tropical Medicine, London, UK.

August 31, 2018

Short title: Robust instrumental variable methods applied to Mendelian randomization.

*Address: Department of Public Health and Primary Care, Strangeways Research Laboratory, 2 Worts Causeway, Cambridge, CB1 8RN, UK. Telephone: +44 1223 748651. Correspondence to: sb452@medschl.cam.ac.uk.

Abstract

Mendelian randomization is the use of genetic variants to make causal inferences from observational data. The field is currently undergoing a revolution fuelled by increasing numbers of genetic variants demonstrated to be associated with exposures in genome-wide association studies, and the public availability of summarized data on genetic associations with exposures and outcomes from large consortia. A Mendelian randomization analysis with many genetic variants can be performed relatively simply using summarized data. However, a causal interpretation is only assured if each genetic variant satisfies the assumptions of an instrumental variable. To provide some protection against failure of these assumptions, robust methods for instrumental variable analysis have been proposed. Here, we develop three extensions to instrumental variable methods using: i) robust regression, ii) the penalization of weights from candidate instruments with heterogeneous causal estimates, and iii) L1 penalization. Results from a wide variety of robust methods, including the recently-proposed MR-Egger and median-based methods, are compared in an extensive simulation study. We demonstrate that two methods, robust regression in an inverse-variance weighted method and a simple median of the causal estimates from the individual variants, have considerably improved Type 1 error rates compared with conventional methods in a wide variety of scenarios when up to 30% of the genetic variants are invalid instruments. While the MR-Egger method gives unbiased estimates when its assumptions are satisfied, these estimates are less efficient than those from other methods and are highly sensitive to violations of the assumptions. Methods that make different assumptions should be used routinely to assess the robustness of findings from applied Mendelian randomization investigations with multiple genetic variants.

Keywords: Mendelian randomization, instrumental variables, robust methods, summarized data, aggregated data, MR-Egger.

1 Introduction

An instrumental variable (denoted Z) is a randomized or quasi-randomized variable that can be used to estimate the causal effect of an exposure (or risk factor, denoted X) on an outcome (Y) in the presence of arbitrary unmeasured confounding [1]. An instrumental variable satisfies three assumptions:

i) association with the exposure: $Z \not\perp\!\!\!\perp X$;

ii) independence from confounders (denoted U) of the exposure–outcome association: $Z \perp\!\!\!\perp U$;

iii) independence from the outcome conditional on the exposure and confounders: $Z \perp\!\!\!\perp Y|X, U$.

The assumptions imply that an instrumental variable cannot have a direct effect on the outcome, but instead any effect is mediated via the exposure (this is known as the exclusion restriction assumption) [2]. A directed acyclic graph illustrating these assumptions is given as Figure 1.

[Figure 1 should appear about here.]

The instrumental variable assumptions are restrictive and often unrealistic in practice. One way of assessing whether these assumptions are satisfied is to compare the causal estimates from several proposed instrumental variables [3]. If the instrumental variable assumptions are satisfied, and under additional parametric assumptions (sufficient conditions are linearity of the instrumental variable–exposure, instrumental variable–outcome and exposure–outcome relationships with no effect heterogeneity), the same causal parameter is estimated by each of the instrumental variables [4]. In this paper, we use the term 'candidate instrument' to describe a variable that is associated with the exposure and hypothesized to satisfy the other instrumental variable assumptions, without prejudicing whether those assumptions are satisfied or not.

A context in which there are often multiple candidate instruments that may plausibly satisfy the instrumental variable assumptions is Mendelian randomization [5, 6], the use of genetic variants as instrumental variables [4, 7]. For complex (i.e. polygenic and multifactorial) exposures, such as body mass index [8] or blood pressure [9], many associated genetic variants have been discovered in genome-wide association studies. A recent development in Mendelian randomization is the availability of summarized data [10]. These data comprise the associations (beta-coefficients and standard errors) of genetic variants with the exposure and with the outcome estimated from univariable regression models. Such associations estimated in large sample sizes have been made publicly available for download by many consortia; examples include associations with glycaemic traits from the Meta-Analyses of Glucose and Insulin-related traits Consortium [11] and with coronary artery disease from the CARDIoGRAMplusC4D consortium [12]. Instrumental variable methods using summarized data have been recently developed, and include an inverse-variance weighted method [13], and two robust methods: a median-based method [14] and MR-Egger [15]; a robust method is

defined here as one that can provide reasonable estimates under weaker assumptions than a conventional approach that assumes all candidate instruments are valid.

In this paper, we consider robust methods for causal inference using multiple instrumental variables, focusing on those that can be implemented using summarized data for uncorrelated candidate instruments. This is typical of an applied Mendelian randomization investigation; when the instrumental variable assumptions are in doubt, it is common to include one genetic variant from each gene region in the analysis. These variants will typically be independently distributed as predicted by Mendel's laws due to their physical separation. Although these robust methods have good theoretical properties, there are several issues particularly with the MR-Egger method, such as low power in realistic scenarios [14], and the influence of outlying variants [16].

In Section 2, we introduce three extensions to existing instrumental variable methods: the use of robust regression, penalization of weights, and L1 penalization. In Section 3, we perform a simulation analysis to compare estimates from various robust methods with respect to bias and coverage properties when some of the candidate instruments do not satisfy the instrumental variable assumptions. Parameters in the simulation analysis are chosen to reflect a typical Mendelian randomization investigation. In Section 4, we show how these methods perform in an applied analysis of the causal effect of body mass index on the risk of schizophrenia. We conclude by discussing the results of this paper, and the potential for future developments (Section 5).

2 Methods

Existing robust methods for causal inference using instrumental variables have taken two different approaches [17]. The first approach (which includes the median-based method [14]) is to assume that some, but not all, of the candidate instruments satisfy the instrumental variable assumptions. The second approach (which includes the MR-Egger method [15]) allows all of the candidate instruments not to be valid instrumental variables, but assumes that the set of variants satisfies a weaker assumption. Software code for implementing all the methods used in this paper is provided in Web Appendix A.1.

2.1 Parametric assumptions

We assume throughout that the parametric assumptions of linearity with no effect heterogeneity hold for the causal exposure–outcome relationship, and for the instrumental variable–exposure and instrumental variable–outcome associations for all valid instruments Z_j, $j = 1, \ldots, J$:

$$\begin{aligned} \mathbb{E}(X|Z_j = z) &= \beta_{X0j} + \beta_{Xj}\, z \\ \mathbb{E}(Y|Z_j = z) &= \beta_{Y0j} + \beta_{Yj}\, z \quad \text{for } j = 1, \ldots, J \\ \mathbb{E}(Y|\,\mathrm{do}(X = x)) &= \theta_0 + \theta\, x \end{aligned} \tag{1}$$

where $\text{do}(X = x)$ is the do-operator of Pearl meaning that the value of the exposure is set to x by intervention [18], and the causal effect parameter $\theta = \dfrac{\beta_{Yj}}{\beta_{Xj}}$ for all valid instruments [4]. Issues relating to the plausibility of these parametric assumptions are left to the discussion.

Under these linearity assumptions, the association of a candidate instrument with the outcome β_{Yj} decomposes into a direct effect α_j and an indirect effect that corresponds to the causal effect of the exposure on the outcome (θ) multiplied by the association of the candidate instrument with the exposure (β_{Xj}) [19]:

$$\beta_{Yj} = \alpha_j + \theta\, \beta_{Xj}. \tag{2}$$

The term 'pleiotropy' refers to a genetic variant having associations with more than one risk factor on different causal pathways [7]. A pleiotropic genetic variant is typically not a valid instrumental variable. In this decomposition, genetic variant j is pleiotropic if $\alpha_j \neq 0$, and α_j is referred to as the pleiotropic effect. The decomposition is illustrated in Figure 2 (the genetic variant is denoted G_j rather than Z_j to emphasize that it is not a valid instrument if $\alpha_j \neq 0$).

The outcome is assumed to be continuous. If the outcome is binary, then the methods can proceed using genetic associations from logistic regression analyses (log odds ratios), provided that the linear assumptions above hold for the logit-transformed probability of the outcome. There are some technical issues with the interpretation of the causal estimate with a binary outcome and a logistic-linear model [20]. However, instrumental variable estimates are typically unbiased under the null in this setting, and Type 1 error rates are not generally inflated (with the notable exception of those from the two-stage residual inclusion method [21]).

[Figure 2 should appear about here.]

2.2 MR-Egger method and the InSIDE assumption

The MR-Egger method is performed by weighted linear regression of the associations of the candidate instruments with the outcome ($\hat{\beta}_{Yj}$) on the associations of the candidate instruments with the exposure ($\hat{\beta}_{Xj}$), using the inverse of the variances of the associations with the outcome ($\text{se}(\hat{\beta}_{Yj})^{-2}$) as weights:

$$\hat{\beta}_{Yj} = \theta_0 + \theta_1\, \hat{\beta}_{Xj} + \epsilon_j, \text{ weights } = \text{se}(\hat{\beta}_{Yj})^{-2}. \tag{3}$$

If the intercept term in this regression model is estimated, then the slope coefficient is the MR-Egger estimate. To ensure that the MR-Egger estimate is invariant to the original arbitrary choice of the coding allele (effect allele) for each genetic variant, we first orientate all the associations so that the $\hat{\beta}_{Xj}$ estimates are all positive [16]. If the intercept term is set to zero, then the slope coefficient is the inverse-variance weighted (IVW) estimate. The IVW estimate can also be obtained by inverse-variance weighted meta-analysis of the ratio estimates $\hat{\theta}_j$ using standard errors $\dfrac{\text{se}(\hat{\beta}_{Yj})}{\hat{\beta}_{Xj}}$. Simulations have shown that this simple choice of standard error expression gives reasonable

inferences in realistic settings [13]. Additionally, the IVW estimate using this expression is the same as the two-stage least squares estimate that can be calculated using individual-level data [22] (assuming, as throughout, that the candidate instruments are uncorrelated).

The ratio estimate based on candidate instrument j is $\dfrac{\hat{\beta}_{Yj}}{\hat{\beta}_{Xj}}$. This is a consistent estimate of the causal effect if $\alpha_j = 0$. As a weighted mean of the ratio estimates, the IVW estimate is consistent if $\alpha_j = 0$ for all j; that is, if none of the candidate instruments are pleiotropic. The MR-Egger method gives consistent estimates of the causal effect under the condition that the pleiotropic effects of the candidate instruments α_j are uncorrelated with the associations of the candidate instruments with the exposure β_{Xj} [15]. This is referred to as the InSIDE assumption (INstrument Strength Independent of Direct Effect). Specifically, we require the weighted covariance $\mathrm{cov}_w(\boldsymbol{\alpha}, \boldsymbol{\beta}_X)$ to be zero:

$$\mathrm{cov}_w(\boldsymbol{\alpha}, \boldsymbol{\beta}_X) = \frac{\sum_j (\alpha_j - \bar{\alpha}_w)(\beta_{Xj} - \bar{\beta}_{Xw})\,\mathrm{se}(\hat{\beta}_{Yj})^{-2}}{\sum_j \mathrm{se}(\hat{\beta}_{Yj})^{-2}} = 0 \tag{4}$$

where $\bar{\beta}_{Xw}$ is the weighted mean of the β_{Xj}, $\bar{\alpha}_w$ is the weighted mean of the α_j, and bold symbols represent vectors across the candidate instruments.

If the intercept term in the MR-Egger analysis differs from zero, then either the InSIDE assumption is violated, or the average pleiotropic effect differs from zero (referred to as directional pleiotropy). In either case, the instrumental variable assumptions are violated for at least one of the candidate instruments, and the IVW estimate will not be consistent. However, provided that the InSIDE assumption holds, the MR-Egger estimate will still be consistent for the causal effect even in the case of directional pleiotropy. The statistical test of whether the intercept term in MR-Egger differs from zero is referred to as the MR-Egger intercept test.

2.3 Motivation: robustness to heterogeneity in causal estimates

The MR-Egger regression has a 100% breakdown point in the sense that all of the candidate instruments can violate the instrumental variable assumptions by having direct effects on the outcome ($\alpha_j \neq 0$), provided that the InSIDE assumption is satisfied. However, it relies on the InSIDE assumption being satisfied for the complete set of candidate instruments. In contrast, in the simple median method (calculated as the median of the ratio estimates from each candidate instrument individually), up to 50% of candidate instruments can violate the instrumental variable assumptions arbitrarily [14]. It would be worthwhile to develop a method that gives robust estimates if either of these two assumptions holds. We propose three novel extensions to existing instrumental variable methods to explore this possibility: 1) robust regression, 2) penalization of weights, and 3) L1 penalization. All approaches downweight the contribution to the causal estimate of candidate instruments with heterogeneous ratio estimates.

2.4 Robust regression

Several methods have been proposed for performing robust regression (that is, regression with greater than a 0% breakdown point) [23]. Here, we use an MM-estimation approach as described by Koller and Stahel [24]. Each of the letter Ms refers to a "maximum likelihood type" maximization step. Briefly, the method proceeds by finding a robust S-estimate ("scale-type estimate") that minimizes an M-estimate of the scale of the residuals (the first M in the method's name). The estimated scale is then held constant whilst a close-by M-estimate of the parameters is located (the second M) [25]. This provides robustness both to outliers and to data points with high leverage. Further robustness is provided by using Tukey's bisquare objective function [26]: instead of minimizing the sum of squared residuals, we minimize the sum of a function of the residuals that is capped at a maximum value for each residual. This means that an outlier in the regression analysis has the same contribution to the objective function no matter how extreme the outlier is.

If the objective function of the standardized residuals r_j in the regression is $\sum_j \rho(r_j)$, then ordinary least squares regression minimizes the sum of the square of the residuals, $\rho(r_j) = r_j^2$. In Tukey's bisquare objective function,

$$\rho(r_j) = \begin{cases} \frac{c^2}{6}\left\{1 - \left[1 - \left(\frac{r_j}{c}\right)^2\right]^3\right\} & \text{if } |r_j| < c \\ \frac{c^2}{6} & \text{if } |r_j| \geq c. \end{cases} \tag{5}$$

The value of the tuning parameter c is chosen as 1.548 to provide a high breakdown point in the S-estimation step, and as 4.685 to provide an efficient estimator in the M-estimation steps. This method for robust regression is the default choice implemented by the `lmrob` command in the R package *robustbase* [27]. Robust regression can replace standard regression in both the IVW and MR-Egger methods.

2.5 Penalization of weights

Another way of providing additional robustness is to penalize the weights of candidate instruments with heterogeneous ratio estimates in the weighted regression model. This could be achieved in many ways; we propose an approach using Cochran's Q statistic as a measure of heterogeneity. For the IVW method:

$$Q = \sum_j Q_j = \sum_j \sigma_{Yj}^{-2}(\hat{\beta}_{Yj} - \hat{\theta}\hat{\beta}_{Xj})^2 \tag{6}$$

where $\hat{\theta}$ is here taken as the IVW estimate. The Q statistic has an approximate χ^2_{J-1} distribution under the null hypothesis that all candidate instruments satisfy the instrumental variable assumptions; the components of the Q statistic for each candidate instrument (Q_j) approximately have χ^2_1 distributions. So as not to distort the majority of weights, we propose penalization using the one-sided upper p-value (denoted q_j) on a χ^2_1 distribution corresponding to Q_j, by multiplying the weight ($\text{se}(\hat{\beta}_{Yj})^{-2}$) by $\min(1, 20q_j)$. The same downweighting factor has previously been

used for weights in the median-based method to give a penalized weighted median estimate [14]. For the median-based methods, we replace the IVW estimate by the relevant median estimate in the calculation of the Q statistic.

For the MR-Egger method, we consider a Q statistic equivalent to the residual sum of squares from the weighted regression, which has an approximate χ^2_{J-2} distribution if the MR-Egger regression model is correct [28]:

$$Q = \sum_j Q_j = \sum_j \sigma_{Yj}^{-2}(\hat{\beta}_{Yj} - \hat{\theta}_0 - \hat{\theta}_1\hat{\beta}_{Xj})^2. \tag{7}$$

If most candidate instruments are valid instrumental variables, then robust regression in either the IVW or MR-Egger method should give a consistent causal estimate asymptotically as the sample size increases, as the association estimates from valid instruments should all align on a straight-line through the origin [19] (the value of "most" depends on the breakdown point of the robust regression method). Equally, provided that the causal estimate in the Q statistic is close to the true causal effect, penalization of weights should downweight the contribution of invalid instruments to zero asymptotically as the sample size increases. If the pleiotropic effects of candidate instruments are independently distributed from their associations with the exposure (a population version of the InSIDE assumption), then the weighted correlation between the pleiotropic effects and associations with the exposure should tend to zero asymptotically as the number of candidate instruments increases for all choices of weights, and should be zero on average for a random choice of candidate instruments. This means that penalization of weights should not affect the consistency of the MR-Egger method under the population InSIDE assumption, nor the use of any robust regression method that is equivalent to varying the weights. This provides some motivation that these extensions should provide reasonable estimates in large samples. However, it is unclear what will happen if some variants satisfy the InSIDE assumption, but others do not.

2.6 L1 penalization

An alternative estimation method is to fit a separate intercept coefficient for each candidate instrument, and to use penalization to identify the model. In the standard MR-Egger method, a single intercept term is estimated, representing the average pleiotropic effect. Weighted linear regression in the MR-Egger method minimizes the following expression:

$$\sum_j \mathrm{se}(\hat{\beta}_{Yj})^{-2}(\hat{\beta}_{Yj} - \theta_0 - \theta_1\hat{\beta}_{Xj})^2 \tag{8}$$

We propose replacing the θ_0 with θ_{0j}, and adding an L1-penalty term:

$$\sum_j \mathrm{se}(\hat{\beta}_{Yj})^{-2}(\hat{\beta}_{Yj} - \theta_{0j} - \theta_1\hat{\beta}_{Xj})^2 + \lambda\sum_j |\theta_{0j}| \tag{9}$$

where λ is a tuning parameter. If $\lambda = 0$, then all the candidate instruments are allowed to be pleiotropic, and the model is not identified. If $\lambda = \infty$, then this is

equivalent to the IVW method using all candidate instruments, as the pleiotropic effects are forced to take the value zero – in effect, all candidate instruments are assumed to be valid instruments. As the value of λ decreases, the number of candidate instruments for which $\alpha_j \neq 0$ increases, and these candidate instruments are allowed to be pleiotropic. An advantage of L1 penalization over other penalization options is the sparsity property – some coefficients are shrunk to zero, representing invalid instruments. Additionally, it can be shown that the estimate of θ_1 in L2 penalization does not depend on the value of the tuning parameter (see Web Appendix A.2).

Once the value of λ is determined, we perform the IVW method using all candidate instruments that are determined to be valid instruments (that is, for all j such that $\hat{\theta}_{0j} = 0$). This provides a causal estimate and a standard error.

3 Simulation study

We perform a simulation study to compare the bias and coverage properties of estimates from different methods:

- standard linear regression without and with an intercept term using inverse-variance weights as in equation (3) – this is equivalent to the IVW and MR-Egger methods respectively;
- robust linear regression (MM-estimation with bisquare objective function) without and with an intercept term using inverse-variance weights;
- standard linear regression without and with an intercept term using penalized inverse-variance weights;
- robust linear regression without and with an intercept term using penalized inverse-variance weights;
- L1 penalization using various approaches for selecting the tuning parameter;
- simple, weighted, and penalized weighted median estimates (for comparison).

We investigate whether these methods give reasonable inferences (in particular, maintain nominal Type 1 error rates under the causal null hypothesis [$\theta = 0$], but have reasonable power under the alternative) in realistic scenarios. Robust regression is implemented using the `lmrob` command from the *robustbase* package in R [29] with the `method = "MM"` option [27].

3.1 Data-generating model

The data-generating model for the simulation study is as follows:

$$
\begin{aligned}
U_i &= \sum_{j=1}^{J} \phi_j G_{ij} + \epsilon_{Ui} \\
X_i &= \sum_{j=1}^{J} \gamma_j G_{ij} + U_i + \epsilon_{Xi} \\
Y_i &= \sum_{j=1}^{J} \alpha_j G_{ij} + \theta X_i + U_i + \epsilon_{Yi} \\
G_{ij} &\sim \text{Binomial}(2, 0.3) \text{ independently for all } j = 1, \ldots, J \\
\epsilon_{Ui}, \epsilon_{Xi}, \epsilon_{Yi} &\sim \mathcal{N}(0, 1) \text{ independently}
\end{aligned}
\tag{10}
$$

for participants indexed by $i = 1, \ldots, N$, and candidate instruments indexed by $j = 1, \ldots, J$. The candidate instruments G_j are simulated to be equivalent to genetic variants that are single nucleotide polymorphisms in Hardy–Weinberg equilibrium with minor allele frequency 0.3. The variable U is a confounder in the relationship between the exposure and the outcome, and is assumed to be unmeasured. The error terms ϵ_{Ui}, ϵ_{Xi}, and ϵ_{Yi} were each drawn independently from standard normal distributions. The causal effect of the exposure on the outcome was either $\theta = 0$ (null causal effect) or $\theta = 0.1$ (positive causal effect). The effects of the candidate instruments on the exposure (γ_j) were drawn from a uniform distribution between 0.03 and 0.1. The direct effects of a candidate instrument (genetic variant) on the outcome (α_j) and the effects of the candidate instruments on the confounder (ϕ_j) were set to zero if the candidate instrument was a valid instrumental variable; for candidate instruments that were invalid instrumental variables:

- In Scenario 2 (direct effects average to zero – balanced pleiotropy, population InSIDE satisfied), the α_j parameters were drawn from a uniform distribution between -0.1 and 0.1, and the ϕ_j were taken as 0.
- In Scenario 3 (direct effects do not average to zero – directional pleiotropy, population InSIDE satisfied), the α_j parameters were drawn from a uniform distribution between 0 and 0.1, and the ϕ_j were taken as 0.
- In Scenario 4 (direct effects operate via confounder and hence do not average to zero – directional pleiotropy, InSIDE not satisfied), the ϕ_j parameters were drawn from a uniform distribution between -0.1 and 0.1, and the α_j were taken as 0.

In Scenario 1, all candidate instruments are taken to be valid instruments. In Scenarios 2 to 4, each candidate instrument was determined to be a valid or an invalid instrumental variable based on a Bernoulli trial with the probability of being invalid set to 0.1, 0.2, and 0.3. Although we only consider scenarios with (on average) up to 30% invalid instruments, the pleiotropic effects of invalid instruments are fairly large.

The maximal indirect association of a candidate instrument with the outcome via the exposure with a positive causal effect is $0.1 \times 0.1 = 0.01$ (if $\gamma_j = 0.1$), whereas the maximal direct (pleiotropic) effect is 0.1 (if either $\alpha_j = 0.1$ or $\phi_j = 0.1$),

A total of 10 000 simulated datasets were generated for $N = 20\,000$ participants and $J = 25$ candidate instruments. A 'two-sample' setting was assumed in which associations of the candidate instruments with the exposure were estimated in N participants, and associations with the outcome in a separate sample of N participants. Results obtained in a one-sample setting in which the associations with the exposure and with the outcome are obtained in the same individuals are given in Web Appendix A.3. Only the summarized data, that is the estimated univariable associations of the candidate instruments with the exposure and with the outcome, and their standard errors, were used by the analysis methods. The average proportion of variance in the exposure explained by the candidate instruments (R^2 statistic) was 2.5% (2.8% in Scenario 4), and the average F statistic was 20.5 (23.3 in Scenario 4).

Six strategies were undertaken for choosing the value of the tuning parameter λ in the L1 penalization method. We set $\lambda = 1$, $\lambda = 2$, and $\lambda = 3$, to compare the performance of the method with different choices of the tuning parameter. Fourthly, we used leave-one-out cross-validation of the likelihood function. Fifthly, we used a grid search to pick the value of λ that gave the causal estimate closest to zero; estimates were calculated for $\lambda = 0.1, 0.2, \ldots, 4.9, 5.0, 5.2, 5.4, \ldots, 9.8, 10.0$. Finally, we used a heterogeneity criterion to determine how many variants to include in the model; we increased the value of the tuning parameter by the same increments as in the grid search, stopping when the residual standard error in the regression model was above 1 and the next value of λ increased the residual standard error by an increment of more than $\chi^2_1(0.95)/(J_{inc} - 1)$, where $\chi^2_1(0.95)$ is the upper 95th percentile of a chi-squared distribution on 1 degree of freedom, and J_{inc} is the number of candidate instruments included in the model. This is motivated by Cochran's Q heterogeneity statistic (equal to the residual standard error multiplied by the number of candidate instruments less 1) having a χ^2_{J-1} distribution under the null hypothesis that all candidate instruments are estimating the same causal parameter.

3.2 Results

Results from the simulation study are provided in Table 1 (Scenario 1 only, all methods), Table 2 (Scenarios 2-4, weights not penalized), Table 3 (Scenarios 2-4, penalized weights), and Table 4 (Scenarios 2-4, L1 penalization methods). Table 1 displays the mean estimate across simulations, standard deviation of estimates, mean standard error of estimates, and the empirical power to detect a causal effect (the proportion of simulations where the 95% confidence interval [estimate $\pm$ 1.96 standard errors] excluded the null). With a null causal effect, power to detect a causal effect is the same as the Type 1 error rate, and the expected power is 5%. In Tables 2, 3 and 4, the mean standard error of estimates is omitted (the pattern of results for the mean standard error was similar to that in Scenario 1 except as noted below). In some of the simulations, the robust regression method did not report a standard error (less than 1% in all cases, except up to 2.5% for the robust method with an intercept in Scenario 4); the number of simulations that failed to report a standard error is given

in Table 1 for Scenario 1, and in Web Table A1 for other scenarios. Simulations were not excluded from the results if a standard error was not reported (except for the calculation of the mean standard error); power calculations include these simulations as if the standard error estimate is infinite. The Monte Carlo standard error (the uncertainty due to the limited number of simulations considered) for the power was 0.2% with a null effect, and between 0.2% and 0.5% with a positive causal effect. A graph illustrating the coverage rates for a limited selection of methods is provided as Figure 3.

Scenario 1 (Table 1): When all candidate instruments were valid instruments, all methods provided unbiased estimates under the null, with Type 1 error rates close to or below the nominal significance level of 5%. The standard deviation of estimates was slightly below the mean standard error of estimates for all methods, with differences most marked for the median-based methods. This difference suggests that methods may be slightly conservative in their inferences. In terms of precision of the causal estimate, regression methods without an intercept (including the IVW method) and L1 penalization methods were the most precise, followed closely by the median-based methods, while regression methods with an intercept (including the MR-Egger method) were the least precise. Differences in precision between the standard, robust and penalized methods were slight. The exception was the L1 penalization method taking the minimal estimate, which gave conservative inference and less variable estimates, particularly under the null.

With a positive causal effect, differing precisions of the causal estimate were also evidenced by the marked differences in power to detect a causal effect. The power for the regression methods including an intercept term was barely above 5%. While precision of the causal estimate for the IVW method depends on the proportion of variance in the exposure explained by the candidate instruments, precision of the causal estimate for MR-Egger depends on the variability between the instrument–exposure associations [30]. If all candidate instruments have exactly the same magnitude of association with the exposure, then the MR-Egger estimate is undefined. The MR-Egger estimate will always be less precise than the IVW estimate, but the difference in precision will depend on whether the instrument–exposure associations for different candidate instruments are similar to each other or not.

While there was some attenuation towards the null with a positive causal effect for all the methods (except for the simple median method) due to uncertainty in the associations of the candidate instruments with the exposure, this was minimal for the IVW and other methods with no intercept, but substantial for the MR-Egger and other methods with an intercept. This attenuation is a known phenomenon called finite-sample bias (also known as weak instrument bias [31]). Bias in the two-sample setting acts towards the null [32] and is related to regression dilution bias [33]; it arises due to measurement error in the independent variable in a regression model. Relative bias of the IVW estimate is around $1/F$, where F is the expected value of the F statistic from regression of the exposure on the IVs (here, $1/F \approx 1/20 = 5\%$, similar to the observed attenuation in the mean IVW estimates) [34]; whereas attenuation of the MR-Egger estimate is approximately equal to the I^2 statistic from meta-analysis of the weighted associations with the exposure $\hat{\beta}_{Xj}\,\mathrm{se}(\hat{\beta}_{Yj})^{-1}$ with standard errors $\mathrm{se}(\hat{\beta}_{Xj})\,\mathrm{se}(\hat{\beta}_{Yj})^{-1}$

[30]. The I^2 statistic is large when the candidate instruments have a wide spread of associations with the exposure or their associations are precisely estimated, and small when their associations with the exposure are imprecisely measured or all similar. In the simulation, the average value of the I^2 statistic was 60.1%. This bias can be corrected using the Simulation Extrapolation (SIMEX) method [35, 30], although this was not computationally feasible in the simulation setting. While measurement error in the exposure can lead to inflation of the intercept term in the MR-Egger method, in this case the 95% confidence interval for the intercept term excluded zero for MR-Egger in 4.7% of simulations – close to the expected nominal 5% level, indicating that over-rejection of the null hypothesis for the MR-Egger intercept test was not evident in this example.

[Table 1 should appear about here.]

Scenario 2, 3 and 4, non-penalized weights (Table 2): Mean estimates in Scenario 2 (balanced pleiotropy, InSIDE satisfied) were unbiased with a null causal effect for all methods. With a positive causal effect, mean estimates were similar to those in Scenario 1: close to unbiased for most methods, but with severe attenuation for regression methods with an intercept term. However, there were marked differences in the precision of estimates compared with Scenario 1. Out of previously proposed methods, estimates from the median-based methods were more precise than those from the IVW method, although this did not translate into greater power with a positive causal effect when only 10% of candidate instruments were not valid instruments. However, the greatest power was obtained by the robust regression method with no intercept. Although the power of the MR-Egger method and other regression methods with an intercept was low, the use of robust regression did reduce the standard deviation and mean standard error of estimates.

Scenario 3 (directional pleiotropy, InSIDE satisfied) demonstrates the value of the MR-Egger and related methods estimating an intercept for providing robust inferences under the InSIDE assumption. While estimates from other methods (particularly the IVW method) were biased under the null, mean estimates from regression methods with an intercept term were close to unbiased, and Type 1 error rates were close to nominal levels. But again, these methods were unable to identify the presence of a causal effect with reasonable power, and mean estimates were substantially attenuated. More seriously, in Scenario 4 (directional pleiotropy, InSIDE not satisfied), the MR-Egger method performed much worse than the IVW method, with mean estimates far more biased and larger Type 1 error rates. While there was some improvement using robust regression with an intercept term when there were few invalid instruments, there was still substantial bias and Type 1 error inflation, as well as even less precise estimates compared with the MR-Egger method when there were many invalid instruments. The MR-Egger and related methods are highly sensitive to the validity of the InSIDE assumption. As the InSIDE assumption is not testable, this is a major limitation of these methods.

In contrast, while the median-based methods and robust regression method without an intercept had bias in mean estimates and inflated Type 1 error rates, rates were substantially below those for the IVW method. In particular, Type 1 error

rates were close to 10% or below for the simple median and robust regression without an intercept methods in Scenarios 3 and 4 with up to 20% invalid instruments, and for the simple median method in Scenario 4 with 30% invalid instruments. The weighted median method was particularly poor in Scenario 4; the data-generating mechanism meant that the invalid instruments received more weight in the analysis than the valid instruments as they had greater associations with the exposure (via an additional association with the confounder). The median-based methods and robust regression without an intercept also had reasonable power to detect a causal effect when present.

[Table 2 should appear about here.]

Scenarios 2, 3 and 4, penalized weights (Table 3): The use of penalized weights generally led to more precise causal estimates, and Type 1 error rates were somewhat improved in Scenarios 3 and 4 for the IVW and weighted median methods. However, Type 1 error rates for the penalized methods generally exceeded nominal levels in all scenarios, especially when 20% or more candidate instruments were invalid. A particular cause for concern is the inflated Type 1 error rates in Scenario 2, which did not occur with unpenalized weights. The reason seems to be that the heterogeneity between the estimates from candidate instruments was underestimated, and hence there was underestimation of the uncertainty in the causal estimate. This highlights a danger that penalization of weights can lead to overconfidence in making inferences, as evidence that points in a different direction is downweighted in the analysis. Penalization of weights did not seem to be a worthwhile strategy for controlling Type 1 error rates in this simulation study.

[Table 3 should appear about here.]

Scenarios 2, 3 and 4, L1 penalization method (Table 4): Similarly, although Type 1 error rates for the implementations of the L1 penalization method were improved compared with the IVW method in Scenarios 3 and 4, there was slight overprecision in Scenario 2, and Type 1 error rates were consistently greater than nominal levels. Estimates using the tuning parameter (λ) chosen by cross-validation had the largest Type 1 error rates and the most variable estimates, but did not always have the greatest power to detect a causal effect. This suggests that cross-validation tended to include too few candidate instruments in the causal analysis, and in Scenarios 3 and 4, often chose the wrong candidate instruments. Estimates using the heterogeneity criterion to choose the values of λ produced better inferences than by simply choosing a fixed value of λ in Scenarios 2 and 3, but worse in Scenario 4. When the value of λ was chosen to give the minimal causal estimate, Type 1 error rates were conservative in Scenario 2, and in Scenario 3 with up to 20% invalid instruments. However, power to detect a causal estimate was also considerably lower.

Supplementary analyses (not done yet for L1 penalization): This simulation was repeated in a one-sample setting in which associations of the candidate instruments with the exposure and with the outcome were obtained in the same sample of 20 000 individuals for the methods using non-penalized weights. Results are

displayed in Web Table A2 (Scenario 1) and Web Table A3 (Scenarios 2 to 4). Bias in the direction of the observational association was observed in all methods except for the simple median method (which remained unbiased in Scenarios 1 and 2). However, the bias of the MR-Egger method was greater and more severe than that of the IVW method: in Scenario 1 with a null causal effect, the mean estimates were 0.024 for the IVW method and 0.173 for the MR-Egger method, and Type 1 error rates were 6.8% and 27.2% respectively. The rejection rate of the MR-Egger intercept test was also inflated (23.5% with a null causal effect, 20.3% with a positive causal effect). The one-sample setting is another case where the MR-Egger method performs poorly.

The simulation was also repeated in a two-sample setting with only 10 candidate instruments, to observe whether the robust methods were able to operate well with fewer instruments to detect violations of the instrumental variables assumptions. Results for Scenarios 2 to 4 are presented in Web Table A4. Power to detect a causal effect was generally much lower, but otherwise similar results were observed.

Finally, Table 5 shows the proportion of datasets for the original simulation study rejecting the causal null using both the simple median and robust regression method with no intercept (robust IVW), and the empirical power of the MR-Egger intercept test for detecting directional pleiotropy and/or violations of the InSIDE assumptions. The combination of the simple median and robust IVW methods generally provided conservative inferences, with Type 1 error rates close to or below nominal levels except in Scenario 3 with 30% invalid instruments. This suggests that multiple robust methods could be used as sensitivity analyses in practice to better control Type 1 error rates. The MR-Egger intercept test is a test of directional pleiotropy and/or violation of the InSIDE assumption: as expected, rejection rates were around 5% in Scenarios 1 and 2, and greater in Scenarios 3 and 4. This suggests that, even if the MR-Egger estimate is unreliable, the method may be useful for detecting in which cases the IVW method is likely to be biased.

[Table 5 should appear about here.]

4 Applied example: causal effect of body mass index on schizophrenia risk

As an applied example to illustrate the methods, we considered the causal effect of body mass index (BMI) on schizophrenia risk. Individuals with schizophrenia generally have higher incidence of obesity than the general population [36], although the relationship is thought to arise from the effect of anti-psychotic medicine on BMI (reverse causation) rather than as a causal effect of BMI [37]. We use 97 genetic variants previously demonstrated to be associated with BMI at a genome-wide level of significance by the Genetic Investigation of Anthropometric Traits (GIANT) consortium [38]. Associations with the exposure were taken from univariable linear regression analyses in up to 339 224 European-descent individuals from the GIANT consortium [38]; associations with the outcome were taken from univariable logistic regression analyses in around 9000 European-descent cases and 8000 controls from the Psychiatric Genomics Consortium [39]. The 97 genetic variants explain about 2.7% of the

variance in BMI. Both sets of genetic associations have previously been made publicly available, and the association estimates can be obtained using the PhenoScanner tool at `http://phenoscanner.medschl.cam.ac.uk/`; they are also displayed visually in Figure 4. The graph indicates that there are several genetic variants that are clear outliers in their associations with schizophrenia, suggesting potential pleiotropy. The I^2 statistic for the weighted genetic associations with the exposure was 88.8%, suggesting that attenuation of the MR-Egger and other methods that estimate an intercept should not be severe.

[Figure 4 should appear about here.]

Estimates and 95% confidence intervals are provided in Table 6. Random-effects models were used in all analyses. Each estimate represents the log odds ratio for schizophrenia per 1 standard deviation increase in BMI. Although all estimates are compatible with the null, there is a wide variation in the standard errors of estimates. Using non-penalized weights, a similar pattern of results was seen as in the simulation analyses of Scenario 2: the robust method with no intercept giving the most precise estimate, followed by the median-based methods, with the MR-Egger method far behind. The use of penalized weights led to large improvements in precision for all except the median-based methods, indicating that although penalization of weights did not seem to add robustness to results in the simulation study, it may have a role in improving the precision of results in cases like this where there are genetic variants that are clear outliers. In the IVW method, the use of penalized weights reduced the residual standard error from 2.14 to 1.12, only slightly above the value of 1 that would be expected in the absence of heterogeneity. In an applied setting, the genetic variants that are downweighted in the analysis should be examined for pleiotropy to determine whether their omission from the analysis is reasonable.

L1 penalization estimates for a range of values of the tuning parameter are displayed in Figure 5. Using the heterogeneity criterion, the value of the tuning parameter was $\lambda = 1.9$ and 64 genetic variants were included in the analysis. Using cross-validation, the value of the tuning parameter was much larger at $\lambda = 6.63$ and 95 of the 97 genetic variants were included in the analysis. The heterogeneity criterion almost chose the value of λ corresponding to the most precise causal estimate ($\lambda = 1.8$ gave a slightly more precise estimate), and a more precise estimate than from any other method. However, as can be seen in Figure 5, causal estimates were fairly similar in magnitude whatever value of λ was chosen, even for $\lambda = 0.1$ when only 5 genetic variants were included in the analysis. While in the simulation study, a strategy was required to choose the value of λ, in practice causal estimates can be compared using a range of values of the tuning parameter.

This applied example illustrates that in addition to providing additional confidence in the robustness of findings from a conventional analysis, the methods introduced in this paper have the potential to improve the efficiency of Mendelian randomization estimates.

[Table 6 should appear about here.]

5 Discussion

In this paper, we have introduced three extensions to instrumental variable methods to downweight the influence of candidate instruments with heterogeneous causal estimates, with the aim of providing more robust estimates in Mendelian randomization investigations. A summary of the methods presented in this paper is provided in Table 7.

[Table 7 should appear about here.]

While the robust and the penalized versions of MR-Egger have desirable theoretical properties, in our simulation study neither method was able to reliably detect the presence of a causal effect of moderate size with reasonable power. Additionally, both these and the original MR-Egger method were highly sensitive to violations of the InSIDE assumption. The MR-Egger intercept test was able to detect scenarios in which the IVW method gave biased estimates, although power was moderate at best. The two methods that had the best performance across the range of scenarios considered in terms of Type 1 error rate and power were the robust version of the IVW method and the simple median method. Although Type 1 error rates were inflated over nominal levels for all methods in at least one scenario, improvement over the standard inverse-variance weighted method was considerable.

If alternative parameters or scenarios were chosen in the simulation study, then different results might have been observed. For example, if candidate instruments had substantially different strengths (and validity of the candidate instruments did not depend on instrument strength, as in Scenario 4), then the weighted median method may have been preferable to the simple median method, and the loss of power in the MR-Egger method compared with the IVW method would have been less severe. Alternatively, if simulations were conducted in a scenario where 100% of the candidate instruments were invalid but they satisfied the InSIDE assumption, then the MR-Egger method would have fared better; likewise if the magnitude of the causal effect was greater (hence the MR-Egger method would have had improved power to detect a causal effect), or if the sample size for the genetic associations with the exposure increased (hence the MR-Egger estimates would have been less attenuated). Hence, the conclusion from this work should not be to promote one method for Mendelian randomization analysis to the exclusion of others, but rather to emphasize the need for multiple sensitivity analyses that make different sets of assumptions. The robust version of the IVW method seems to be a worthwhile sensitivity analysis method in addition to other robust methods previously proposed (such as simple and weighted median, and MR-Egger [17]). The use of penalized weights may be worthwhile to improve precision if a small number of candidate instruments have clearly heterogeneous causal estimates (as demonstrated in the applied example), but the approach is unlikely to lead to robust inferences if several candidate instruments are not valid. Similarly, L1 penalization can improve precision when causal estimtes are heterogeneous, and the approach gave improved inferences over the IVW method when the InSIDE assumption was satisfied. While it is not clear how to best choose the tuning parameter in an automated way for a simulation analysis, in an applied example estimates can be reported for a range of values of this parameter. Additionally, if we had

considered weaker pleiotropic effects in the simulation study, Type 1 error inflation would have been less pronounced.

5.1 Linearity and homogeneity assumptions

In the specification of the analysis models, we have assumed linearity and homogeneity (no effect modification) of the causal effect of the exposure on the outcome, and of the associations of the candidate instruments with the exposure and with the outcome. These assumptions are not necessary to identify a causal effect; weaker assumptions can be made [40] (such as monotonicity of the causal effect [41] or a weaker version of the homogeneity assumption for the causal effect [42, 43]). If the linearity and homogeneity assumptions are violated, then the causal estimate using a single instrumental variable is a valid test of the null hypothesis that the exposure does not have a causal effect on the outcome [4]; this also applies to the causal estimate from the IVW method using multiple instruments, as this is a linear combination of the causal estimates from the individual instruments [44]. Hence, even when the linearity and homogeneity assumptions are violated, the methods proposed in this paper can still be used for the assessment of the causal null hypothesis (does the exposure have a causal effect on the outcome?), even if the estimate does not have a literal interpretation [45].

Additionally, while the linearity and homogeneity assumptions are stringent, genetic variants tend to have small effects on the exposure and outcome. This means that linearity and homogeneity may not be unreasonable assumptions in an applied Mendelian randomization investigation. Linearity and homogeneity in the genetic associations are not required across the whole distribution of the exposure and the outcome, but simply in the range of values predicted by the genetic variants.

5.2 Alternative robust methods

Several other methods have been developed for robust estimation using instrumental variables. Kolesár et al. proposed a method within the framework of k-class estimators with a 100% breakdown level under the InSIDE assumption [46]. Kang et al. proposed a method using individual-level data based on penalized regression for detecting and accounting for invalid instruments that provides a consistent estimate of causal effect if at least 50% of the candidate instruments are valid using L1 penalization to downweight the contribution to the analysis of candidate instruments that have heterogeneous causal estimates [19]. Han proposed a similar penalized estimator within the generalized method of moments framework, again with a 50% breakdown level [47]. These ideas were developed further by Windmeijer et al. [48]; several of the choices made here relating to the L1 penalization method (such as the decision to use the method to identify invalid instruments and to obtain the causal estimate using only the valid instruments, referred to by Windmeijer et al. as a 'post-lasso' estimator, and the decision to explore a heterogeneity criterion for selecting the tuning parameter) were guided by that paper. However, each of these methods requires individual-level data, limiting their applicability to applied Mendelian randomization investigations.

5.3 Conclusion

We have shown that it is difficult to find methods that give robust causal inferences with invalid instruments. Even in the examples with moderate numbers of invalid instruments considered in this paper, all methods had inflated Type 1 error rates in at least one scenario. Nevertheless, although the methods we have proposed are far from perfect, they have much improved Type 1 error rates compared with the conventional IVW method and the recently introduced MR-Egger method in scenarios where the InSIDE assumption fails to hold.

We have demonstrated that using multiple methods for instrumental variable analysis (particularly methods that provide consistent estimates under different assumptions) can provide more reliable inferences for Mendelian randomization investigations. A causal conclusion is more plausible in cases where multiple methods suggest a causal effect. We suggest that the IVW method using robust regression is a worthwhile method to apply in addition to previously proposed methods (in particular the simple median method), and that the use of penalized weights and L1 penalization may be valuable in some situations.

Acknowledgements

Stephen Burgess is supported by the Wellcome Trust (grant number 100114). Jack Bowden is supported by a Methodology Research Fellowship from the Medical Research Council (grant number MR/N501906/1). Frank Dudbridge is supported by the Medical Research Council (grant number K006215). Simon G. Thompson is supported by the British Heart Foundation (grant number CH/12/2/29428).

References

[1] Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 2000; **29**(4):722–729, doi:10.1093/ije/29.4.722.

[2] Clarke PS, Windmeijer F. Instrumental variable estimators for binary outcomes. *Journal of the American Statistical Association* 2012; **107**(500):1638–1652, doi:10.1080/01621459.2012.734171.

[3] Small DS. Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *Journal of the American Statistical Association* 2007; **102**(479):1049–1058, doi:10.1198/016214507000000608.

[4] Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research* 2007; **16**(4):309–330, doi:10.1177/0962280206077743.

[5] Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 2003; **32**(1):1–22, doi:10.1093/ije/dyg070.

[6] Burgess S, Thompson SG. *Mendelian randomization: methods for using genetic variants in causal estimation*. Chapman & Hall, 2015.

[7] Lawlor D, Harbord R, Sterne J, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine* 2008; **27**(8):1133–1163, doi:10.1002/sim.3034.

[8] Speliotes E, Willer C, Berndt S, Monda K, Thorleifsson G, Jackson A, Allen HL, Lindgren C, Luan J, Mägi R, *et al.*. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics* 2010; **42**(11):937–948, doi:10.1038/ng.686.

[9] The International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011; **478**:103–109, doi:10.1038/nature10405.

[10] Burgess S, Scott R, Timpson N, Davey Smith G, Thompson SG, EPIC-InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *European Journal of Epidemiology* 2015; **30**(7):543–552, doi:10.1007/s10654-015-0011-z.

[11] Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, Mägi R, Strawbridge RJ, Rehnberg E, Gustafsson S, *et al.*. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genetics* 2012; **44**(9):991–1005, doi:10.1038/ng.2385.

[12] CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics* 2015; **47**:1121–1130, doi:10.1038/ng.3396.

[13] Burgess S, Butterworth AS, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology* 2013; **37**(7):658–665, doi:10.1002/gepi.21758.

[14] Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology* 2016; **40**(4):304–314, doi:10.1002/gepi.21965.

[15] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* 2015; **44**(2):512–525.

[16] Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Under review.

[17] Burgess S, Bowden J, Fall T, Ingelsson E, Thompson S. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiology* 2016; Accepted manuscript; awaiting proofs.

[18] Pearl J. *Causality: models, reasoning, and inference*. Cambridge University Press, 2000.

[19] Kang H, Zhang A, Cai T, Small D. Instrumental variables estimation with some invalid instruments, and its application to Mendelian randomisation. *Journal of the American Statistical Association* 2015; doi:10.1080/01621459.2014.994705.

[20] Burgess S, CHD CRP Genetics Collaboration. Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. *Statistics in Medicine* 2013; **32**(27):4726–4747, doi:10.1002/sim.5871.

[21] Vansteelandt S, Bowden J, Babanezhad M, Goetghebeur E. On instrumental variables estimation of causal odds ratios. *Statistical Science* 2011; **26**(3):403–422, doi:10.1214/11-sts360.

[22] Burgess S, Bowden J. Integrating summarized data from multiple genetic variants in Mendelian randomization: bias and coverage properties of inverse-variance weighted methods. *arXiv* 2015; **1512.04486**. Available at http://arxiv.org/abs/1512.04486.

[23] Huber PJ. *Robust statistics*. Springer, 2011.

[24] Koller M, Stahel WA. Sharpening wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis* 2011; **55**(8):2504–2515, doi:10.1016/j.csda.2011.02.014.

[25] Yohai VJ. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics* 1987; **15**(20):642–656, doi:10.1214/aos/1176350366.

[26] Mosteller F, Tukey JW. *Data analysis and regression: a second course in statistics*. Addison-Wesley Series in Behavioral Science: Quantitative Methods, 1977.

[27] Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, Koller M, Maechler M. *robustbase: Basic Robust Statistics* 2015. URL `http://cran.r-project.org/package=robustbase`, R package version 0.92-5.

[28] Rücker G, Schwarzer G, Carpenter JR, Binder H, Schumacher M. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics* 2011; **12**(1):122–142.

[29] R Core Team. *R: A Language and Environment for Statistical Computing. Version 3.1.2 (Pumpkin Helmet)*. R Foundation for Statistical Computing, Vienna, Austria 2014. URL `http://www.R-project.org/`.

[30] Bowden J, Del Greco F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for Mendelian randomization analyses using MR-Egger regression: the role of the I^2 statistic 2016. Under review.

[31] Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Statistics in Medicine* 2011; **30**(11):1312–1323, doi:10.1002/sim.4197.

[32] Pierce B, Burgess S. Efficient design for Mendelian randomization studies: subsample and two-sample instrumental variable estimators. *American Journal of Epidemiology* 2013; **178**(7):1177–1184, doi:10.1093/aje/kwt084.

[33] Frost C, Thompson S. Correcting for regression dilution bias: comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2000; **163**(2):173–189, doi:10.1111/1467-985x.00164.

[34] Staiger D, Stock J. Instrumental variables regression with weak instruments. *Econometrica* 1997; **65**(3):557–586.

[35] Cook JR, Stefanski LA. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* 1994; **89**(428):1314–1328, doi:10.1080/01621459.1994.10476871.

[36] Coodin S. Body mass index in persons with schizophrenia. *Canadian Journal of Psychiatry* 2001; **46**(6):549–555.

[37] Allison DB, Fontaine KR, Heo M, Mentore JL, Cappelleri JC, Chandler LP, Cheskin LJ. The distribution of body mass index among individuals with and without schizophrenia. *The Journal of Clinical Psychiatry* 1999; **60**(4):215–220, doi:10.4088/jcp.v60n0402.

[38] Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, *et al.*. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**(7538):197–206, doi:10.1038/nature14177.

[39] Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013; **381**:1371–1379, doi:10.1016/s0140-6736(12)62129-1.

[40] Swanson S, Hernán M. Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology* 2013; **24**(3):370–374, doi:10.1097/ede.0b013e31828d0590.

[41] Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica* 1994; **62**(2):467–475, doi:10.2307/2951620.

[42] Robins JM. *Health service research methodology: a focus on AIDS*, chap. The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. National Center for Health Services Research, 1989; 113–159.

[43] Hernán M, Robins J. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; **17**(4):360–372, doi:10.1097/01.ede.0000222409.00878.37.

[44] Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Statistics in Medicine* 2016; **35**(11):1880–1906, doi: 10.1002/sim.6835.

[45] Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors. *Journal of Clinical Epidemiology* 2015; doi:10.1016/j.jclinepi.2015.08.001.

[46] Kolesár M, Chetty R, Friedman J, Glaeser E, Imbens G. Identification and inference with many invalid instruments. *Journal of Business & Economic Statistics* 2014; doi:10.1080/07350015.2014.978175.

[47] Han C. Detecting invalid instruments using L1-GMM. *Economics Letters* 2008; **101**:285–287.

[48] Windmeijer F, Farbmacher H, Davies N, Davey Smith G. On the use of the lasso for instrumental variables estimation with some invalid instruments. *Technical Report Discussion Paper 16/674*, University of Bristol 2016.

[49] Thompson S, Sharp S. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**(20):2693–2708.

Tables

Method	Scenario 1 Mean	SD	Mean SE	Power	NA [1]
Null causal effect: $\theta = 0$					
Standard, no intercept [2]	0.000	0.044	0.047	3.9	-
Standard, intercept [3]	0.002	0.125	0.133	3.8	-
Robust, no intercept	0.000	0.046	0.050	4.5	1
Robust, intercept	0.002	0.130	0.141	5.7	4
Penalized standard, no intercept	0.000	0.046	0.046	5.2	-
Penalized standard, intercept	0.002	0.130	0.130	4.8	-
Penalized robust, no intercept	0.000	0.047	0.047	5.9	2
Penalized robust, intercept	0.001	0.132	0.134	7.1	5
Simple median	0.000	0.059	0.070	1.8	-
Weighted median	0.001	0.056	0.064	2.1	-
Penalized weighted median	0.001	0.059	0.064	3.1	-
L1 penalization, $\lambda = 1$	0.000	0.055	0.053	5.7	-
L1 penalization, $\lambda = 2$	0.001	0.049	0.046	6.4	-
L1 penalization, $\lambda = 3$	0.000	0.045	0.047	4.2	-
L1 penalization, cross-validation	0.001	0.047	0.046	5.4	-
L1 penalization, minimal estimate	0.000	0.029	0.062	0.9	-
L1 penalization, heterogeneity	0.000	0.045	0.047	4.0	-
Positive causal effect: $\theta = +0.1$					
Standard, no intercept [2]	0.096	0.047	0.050	49.3	-
Standard, intercept [3]	0.065	0.136	0.141	6.7	-
Robust, no intercept	0.096	0.048	0.052	46.1	2
Robust, intercept	0.064	0.140	0.148	8.7	4
Penalized standard, no intercept	0.096	0.049	0.048	51.8	-
Penalized standard, intercept	0.064	0.140	0.137	8.3	-
Penalized robust, no intercept	0.096	0.050	0.050	50.1	2
Penalized robust, intercept	0.064	0.142	0.140	10.3	5
Simple median	0.101	0.063	0.074	24.7	-
Weighted median	0.093	0.059	0.067	25.7	-
Penalized weighted median	0.093	0.062	0.067	26.5	-
L1 penalization, $\lambda = 1$	0.097	0.059	0.056	40.8	-
L1 penalization, $\lambda = 2$	0.097	0.052	0.048	51.9	-
L1 penalization, $\lambda = 3$	0.096	0.047	0.049	50.0	-
L1 penalization, cross-validation	0.096	0.051	0.049	50.4	-
L1 penalization, minimal estimate	0.064	0.048	0.067	19.2	-
L1 penalization, heterogeneity	0.096	0.048	0.049	48.9	-

Table 1: Mean, standard deviation (SD), mean standard error (mean SE) of estimates, and empirical power (%) from weighted linear regression models (weights are penalized where indicated) using standard and robust regression, without and with an intercept term, and median-based methods for Scenario 1.

[1] Number of the 10 000 simulations that failed to report a standard error.

[2] This is the standard inverse-variance weighted (IVW) method, equivalent to the two-stage least squares (2SLS) method with individual-level data.

[3] This is the MR-Egger method.

Method	Scenario 2			Scenario 3			Scenario 4		
	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power
Null causal effect: $\theta = 0$									
Proportion of invalid instrumental variables: 0.1									
Standard, no intercept	0.000	0.069	5.4	0.067	0.067	13.0	0.059	0.076	19.0
Standard, intercept	0.001	0.197	5.6	0.002	0.192	5.8	0.148	0.245	30.7
Robust, no intercept	0.000	0.052	5.1	0.022	0.054	6.1	0.017	0.058	5.6
Robust, intercept	0.001	0.153	6.8	0.001	0.154	6.5	0.081	0.205	11.5
Simple median	0.000	0.066	2.8	0.030	0.066	4.5	0.013	0.066	3.2
Weighted median	0.000	0.061	3.1	0.024	0.061	4.3	0.040	0.077	10.6
Proportion of invalid instrumental variables: 0.2									
Standard, no intercept	-0.001	0.087	6.1	0.135	0.084	34.8	0.112	0.087	34.0
Standard, intercept	0.003	0.251	6.2	0.008	0.234	6.1	0.255	0.260	43.9
Robust, no intercept	0.000	0.064	5.4	0.062	0.074	10.4	0.049	0.080	9.7
Robust, intercept	0.003	0.192	7.1	0.006	0.192	7.0	0.207	0.269	24.9
Simple median	-0.001	0.074	3.8	0.067	0.078	11.3	0.027	0.074	4.9
Weighted median	0.000	0.069	4.5	0.054	0.072	11.5	0.088	0.103	26.4
Proportion of invalid instrumental variables: 0.3									
Standard, no intercept	0.001	0.103	5.9	0.204	0.093	59.3	0.161	0.092	48.9
Standard, intercept	0.000	0.288	5.9	0.005	0.263	6.0	0.327	0.256	50.6
Robust, no intercept	0.001	0.082	5.7	0.122	0.100	19.7	0.099	0.102	19.3
Robust, intercept	0.003	0.243	6.3	0.005	0.239	7.5	0.332	0.293	42.0
Simple median	0.000	0.082	5.1	0.115	0.094	25.0	0.046	0.084	8.6
Weighted median	0.001	0.079	6.7	0.094	0.090	24.3	0.148	0.125	46.2
Positive causal effect: $\theta = +0.1$									
Proportion of invalid instrumental variables: 0.1									
Standard, no intercept	0.095	0.070	32.6	0.162	0.069	69.2	0.155	0.078	63.3
Standard, intercept	0.066	0.202	7.1	0.066	0.196	7.0	0.221	0.252	38.0
Robust, no intercept	0.095	0.055	40.6	0.120	0.057	53.8	0.114	0.062	45.2
Robust, intercept	0.066	0.162	9.0	0.066	0.162	8.8	0.149	0.218	15.8
Simple median	0.100	0.070	23.4	0.132	0.070	38.5	0.114	0.070	29.4
Weighted median	0.093	0.064	24.8	0.117	0.064	37.1	0.134	0.081	45.3
Proportion of invalid instrumental variables: 0.2									
Standard, no intercept	0.095	0.089	22.6	0.230	0.085	84.0	0.208	0.089	73.2
Standard, intercept	0.068	0.255	7.3	0.073	0.238	7.0	0.335	0.266	52.0
Robust, no intercept	0.096	0.067	32.5	0.162	0.078	58.5	0.148	0.084	45.6
Robust, intercept	0.069	0.201	9.1	0.071	0.201	8.8	0.278	0.281	29.5
Simple median	0.100	0.077	22.9	0.170	0.083	53.3	0.129	0.078	33.9
Weighted median	0.093	0.072	24.6	0.149	0.077	50.8	0.185	0.107	62.5
Proportion of invalid instrumental variables: 0.3									
Standard, no intercept	0.097	0.105	18.6	0.299	0.095	93.8	0.257	0.094	81.6
Standard, intercept	0.065	0.291	6.5	0.070	0.267	6.9	0.411	0.261	59.9
Robust, no intercept	0.096	0.085	25.2	0.223	0.102	64.2	0.200	0.104	52.7
Robust, intercept	0.068	0.251	7.7	0.070	0.247	8.8	0.406	0.304	46.7
Simple median	0.101	0.087	22.5	0.221	0.100	69.2	0.148	0.089	38.9
Weighted median	0.094	0.083	24.9	0.191	0.095	65.8	0.245	0.128	76.8

Table 2: Mean, standard deviation (SD) of estimates, and empirical power (%) from weighted linear regression models (weights are not penalized) using standard and robust regression, without and with an intercept term, and simple and weighted median methods for Scenarios 2, 3, and 4. (Note: power with a null causal effect is the Type 1 error rate.)

Method	Scenario 2			Scenario 3			Scenario 4		
	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power
Null causal effect: $\theta = 0$									
Proportion of invalid instrumental variables: 0.1									
Penalized standard, no intercept	0.000	0.051	6.8	0.022	0.053	9.5	0.019	0.056	11.3
Penalized standard, intercept	0.001	0.149	7.5	0.001	0.154	8.2	0.093	0.198	22.7
Penalized robust, no intercept	0.000	0.052	6.3	0.018	0.053	8.4	0.015	0.055	8.2
Penalized robust, intercept	0.001	0.151	7.9	0.001	0.153	7.9	0.077	0.191	15.4
Penalized weighted median	0.001	0.063	3.7	0.011	0.063	4.0	0.016	0.074	6.6
Proportion of invalid instrumental variables: 0.2									
Penalized standard, no intercept	0.000	0.059	9.6	0.056	0.068	24.6	0.049	0.074	25.2
Penalized standard, intercept	0.004	0.178	10.9	0.006	0.194	13.5	0.206	0.241	43.4
Penalized robust, no intercept	0.000	0.059	7.0	0.046	0.066	15.8	0.038	0.070	13.2
Penalized robust, intercept	0.003	0.178	9.1	0.005	0.190	11.4	0.177	0.241	28.5
Penalized weighted median	0.000	0.070	5.3	0.026	0.073	7.2	0.050	0.108	16.8
Proportion of invalid instrumental variables: 0.3									
Penalized standard, no intercept	0.001	0.069	13.3	0.106	0.088	46.9	0.091	0.092	43.9
Penalized standard, intercept	0.003	0.210	14.6	0.005	0.247	20.7	0.309	0.259	61.4
Penalized robust, no intercept	0.001	0.068	7.4	0.087	0.085	27.8	0.070	0.089	20.7
Penalized robust, intercept	0.004	0.213	9.4	0.005	0.238	14.3	0.281	0.270	43.1
Penalized weighted median	0.001	0.079	6.4	0.051	0.092	12.9	0.104	0.144	31.7
Positive causal effect: $\theta = +0.1$									
Proportion of invalid instrumental variables: 0.1									
Penalized standard, no intercept	0.095	0.054	49.0	0.119	0.056	65.2	0.116	0.060	61.3
Penalized standard, intercept	0.065	0.158	9.9	0.066	0.162	10.5	0.164	0.209	29.5
Penalized robust, no intercept	0.095	0.055	43.8	0.115	0.056	57.2	0.111	0.059	53.2
Penalized robust, intercept	0.065	0.159	10.5	0.066	0.161	10.8	0.147	0.202	22.0
Penalized weighted median	0.093	0.066	25.8	0.104	0.067	31.1	0.110	0.079	33.6
Proportion of invalid instrumental variables: 0.2									
Penalized standard, no intercept	0.095	0.062	46.8	0.154	0.071	78.4	0.147	0.077	72.1
Penalized standard, intercept	0.069	0.186	13.6	0.070	0.202	15.3	0.282	0.252	51.8
Penalized robust, no intercept	0.096	0.062	37.9	0.145	0.070	65.7	0.136	0.074	57.0
Penalized robust, intercept	0.069	0.187	11.4	0.070	0.198	13.8	0.254	0.253	36.1
Penalized weighted median	0.093	0.074	25.3	0.148	0.078	37.4	0.147	0.113	45.0
Proportion of invalid instrumental variables: 0.3									
Penalized standard, no intercept	0.097	0.073	46.1	0.205	0.091	89.3	0.190	0.096	82.6
Penalized standard, intercept	0.069	0.219	16.9	0.069	0.254	21.6	0.390	0.270	68.8
Penalized robust, no intercept	0.096	0.072	33.1	0.187	0.089	72.8	0.171	0.094	60.1
Penalized robust, intercept	0.069	0.221	12.0	0.070	0.247	16.6	0.363	0.282	50.6
Penalized weighted median	0.093	0.083	25.1	0.148	0.097	46.4	0.203	0.119	58.4

Table 3: Mean, standard deviation (SD) of estimates, and empirical power (%) from weighted linear regression models (weights are penalized) using standard and robust regression, without and with an intercept term, and penalized weighted median method for Scenarios 2, 3, and 4.

Method	Scenario 2			Scenario 3			Scenario 4		
	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power
Null causal effect: $\theta = 0$									
Proportion of invalid instrumental variables: 0.1									
$\lambda = 1$	0.000	0.062	6.9	0.025	0.062	8.7	0.040	0.074	14.0
$\lambda = 2$	0.000	0.053	7.4	0.015	0.055	8.8	0.025	0.066	12.5
$\lambda = 3$	0.000	0.050	5.2	0.012	0.052	6.4	0.022	0.071	11.5
Cross-validation	0.001	0.090	6.0	0.070	0.103	10.9	0.118	0.155	36.0
Minimal estimate	0.000	0.028	1.0	0.012	0.030	1.7	0.017	0.042	3.5
Heterogeneity criterion	0.000	0.062	5.4	0.028	0.058	6.2	0.051	0.092	20.9
Proportion of invalid instrumental variables: 0.2									
$\lambda = 1$	0.000	0.069	8.5	0.057	0.076	17.2	0.104	0.119	31.6
$\lambda = 2$	0.000	0.059	8.4	0.037	0.067	15.2	0.082	0.122	29.7
$\lambda = 3$	0.000	0.058	7.0	0.034	0.066	11.8	0.088	0.136	31.4
Cross-validation	0.001	0.142	8.3	0.192	0.153	33.3	0.273	0.178	70.1
Minimal estimate	0.000	0.030	1.0	0.026	0.041	4.2	0.057	0.092	15.0
Heterogeneity criterion	0.000	0.067	7.0	0.046	0.065	11.0	0.099	0.127	35.8
Proportion of invalid instrumental variables: 0.3									
$\lambda = 1$	-0.001	0.079	10.1	0.104	0.108	30.4	0.206	0.176	55.6
$\lambda = 2$	0.000	0.068	10.3	0.080	0.102	28.5	0.187	0.185	55.5
$\lambda = 3$	0.000	0.068	8.0	0.078	0.103	24.5	0.203	0.189	58.9
Cross-validation	0.001	0.189	10.4	0.321	0.190	59.9	0.389	0.161	88.1
Minimal estimate	0.000	0.034	1.4	0.053	0.072	11.1	0.136	0.151	37.2
Heterogeneity criterion	0.000	0.074	7.7	0.076	0.091	19.7	0.184	0.173	56.5
Positive causal effect: $\theta = +0.1$									
Proportion of invalid instrumental variables: 0.1									
$\lambda = 1$	0.095	0.065	37.5	0.121	0.066	53.2	0.138	0.079	59.4
$\lambda = 2$	0.095	0.056	46.8	0.111	0.058	57.6	0.123	0.071	61.9
$\lambda = 3$	0.095	0.053	45.0	0.109	0.055	54.4	0.121	0.078	57.0
Cross-validation	0.095	0.090	35.5	0.167	0.105	59.3	0.213	0.157	69.6
Minimal estimate	0.053	0.051	12.7	0.082	0.053	28.5	0.091	0.063	34.4
Heterogeneity criterion	0.096	0.066	39.3	0.125	0.062	54.7	0.149	0.096	60.6
Proportion of invalid instrumental variables: 0.2									
$\lambda = 1$	0.095	0.073	36.2	0.156	0.080	64.5	0.206	0.125	75.9
$\lambda = 2$	0.095	0.063	44.2	0.136	0.072	66.2	0.183	0.127	74.5
$\lambda = 3$	0.095	0.061	41.3	0.133	0.071	61.8	0.192	0.142	70.5
Cross-validation	0.094	0.142	23.6	0.287	0.156	75.8	0.370	0.179	87.6
Minimal estimate	0.045	0.054	9.9	0.104	0.063	39.8	0.142	0.109	52.3
Heterogeneity criterion	0.095	0.071	35.7	0.144	0.070	60.2	0.201	0.132	71.9
Proportion of invalid instrumental variables: 0.3									
$\lambda = 1$	0.095	0.083	34.1	0.206	0.113	75.5	0.310	0.179	86.4
$\lambda = 2$	0.095	0.072	41.4	0.180	0.108	75.5	0.293	0.188	86.5
$\lambda = 3$	0.096	0.071	38.3	0.180	0.109	71.4	0.309	0.192	85.0
Cross-validation	0.095	0.190	19.2	0.419	0.192	88.8	0.486	0.162	95.7
Minimal estimate	0.042	0.057	8.6	0.140	0.089	53.5	0.231	0.163	70.8
Heterogeneity criterion	0.096	0.079	33.8	0.176	0.098	69.2	0.288	0.176	83.2

Table 4: Mean, standard deviation (SD) of estimates, and empirical power (%) from L1 penalization methods using different strategies for choosing the tuning parameter λ for Scenarios 2, 3, and 4.

Proportion	Simple median and robust IVW methods				MR-Egger intercept test			
invalid	Scenario 1	2	3	4	Scenario 1	2	3	4
Null causal effect: $\theta = 0$								
0%	0.8%	-	-	-	3.8%	-	-	-
10%	-	1.2%	2.2%	1.5%	-	5.5%	6.4%	24.4%
20%	-	1.6%	5.5%	2.5%	-	6.0%	9.4%	31.2%
30%	-	2.0%	14.1%	6.2%	-	5.9%	13.1%	32.8%
Positive causal effect: $\theta = +0.1$								
0%	21.2%	-	-	-	4.7%	-	-	-
10%	-	19.1%	33.0%	23.6%	-	5.8%	8.0%	21.8%
20%	-	16.7%	44.1%	26.6%	-	6.1%	11.1%	28.2%
30%	-	14.6%	55.8%	31.8%	-	6.0%	15.3%	30.0%

Table 5: Proportion of simulated datasets for which the simple median and robust regression with no intercept (robust IVW) methods rejected the causal null (left), empirical power of the intercept test in MR-Egger method for detecting directional pleiotropy and/or violation of the InSIDE assumption in all scenarios.

	Non-penalized weights		Penalized weights	
Method	Estimate (SE)	95% CI	Estimate (SE)	95% CI
Standard, no intercept	-0.031 (0.100)	-0.227, 0.165	-0.034 (0.057)	-0.147, 0.078
Standard, intercept	0.336 (0.241)	-0.136, 0.808	0.154 (0.143)	-0.127, 0.435
Robust, no intercept	-0.024 (0.079)	-0.180, 0.132	-0.033 (0.062)	-0.154, 0.089
Robust, intercept	0.255 (0.212)	-0.162, 0.671	0.142 (0.150)	-0.152, 0.436
Simple median	-0.073 (0.088)	-0.244, 0.098	-	-
Weighted median	-0.075 (0.087)	-0.246, 0.096	-0.076 (0.090)	-0.253, 0.100
L1 pen, cross-validation	-0.036 (0.087)	-0.207, 0.136	-	-
L1 pen, heterogeneity	-0.022 (0.055)	-0.131, 0.086	-	-

Table 6: Estimates (standard errors, SE) and 95% confidence intervals (CI, calculated as estimate $\pm$ 1.96 standard errors) of causal effect of body mass index on schizophrenia risk (log odds ratio for schizophrenia per 1 standard deviation increase in body mass index).

Method	Description
Inverse-variance weighted (IVW) method	Standard weighted regression with inverse-variance weights and intercept term set to zero.
MR-Egger method	Standard weighted regression with inverse-variance weights and intercept term estimated.
Median-based method	Simple median method is the median of the causal estimates based on the individual candidate instruments. Weighted median method uses inverse-variance weights so that more precise estimates receive more weight in the analysis.
Robust regression (MM-estimation with bisquare objective function)	Standard regression in either the IVW (no intercept) or the MR-Egger (intercept) method can be replaced with robust regression.
Penalization of weights	Inverse-variance weights in either the IVW, MR-Egger, or weighted median method can be replaced with weights that depend on the heterogeneity of the causal estimates – candidate instruments with outlying estimates are downweighted depending on the degree of heterogeneity.
L1 penalization	A separate intercept term, representing the pleiotropic effect of the candidate instrument on the outcome, is allowed for each candidate instrument, but the sum of the absolute values of the pleiotropic effects is not allowed to be too large. The value of the tuning parameter, which regulates the extent to which pleiotropic effects are penalized, must be chosen carefully.

Table 7: Summary of methods investigated in this paper.

Figures

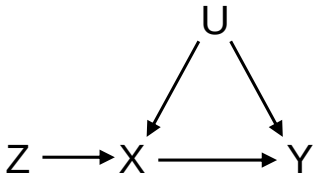


Figure 1: Directed acyclic graph of graphical instrumental variable assumptions.

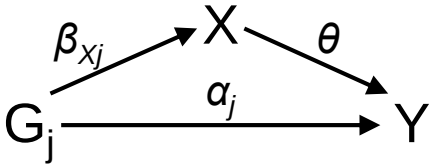


Figure 2: Decomposition of association with the outcome Y for genetic variant G_j into indirect (causal) effect via the exposure X and direct (pleiotropic) effect (see equation 2).

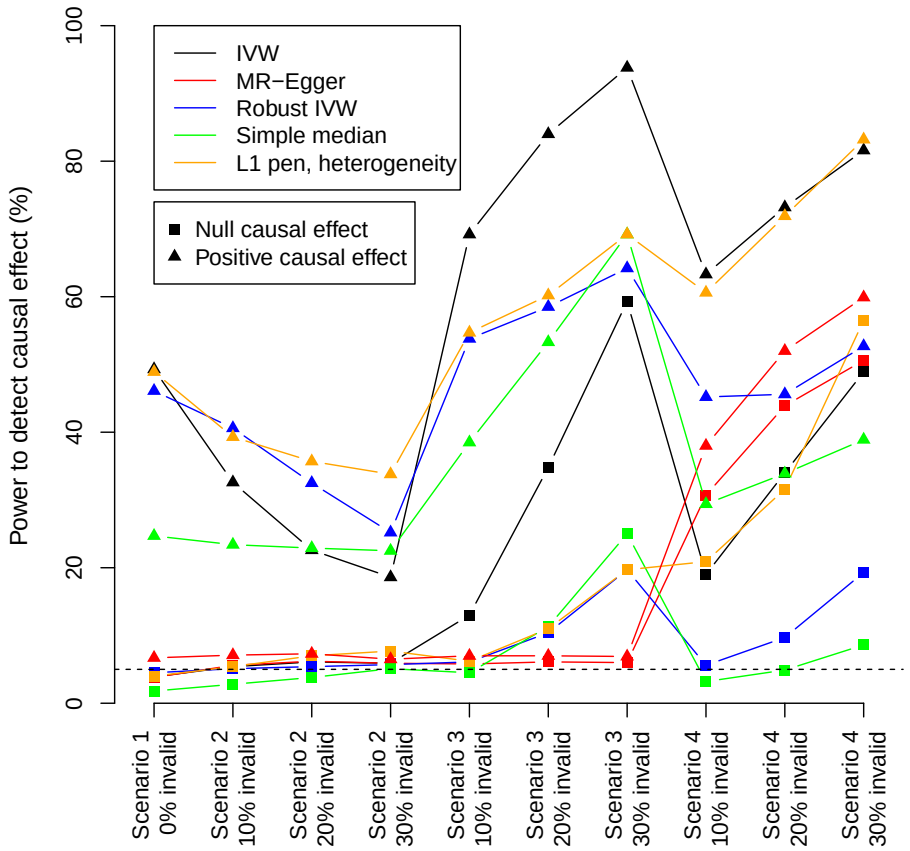


Figure 3: Power to detect a causal effect (equivalent to Type 1 error rate with null causal effect) for selected methods in each scenario. The dashed line is at 5%; the nominal power expected with a null causal effect.

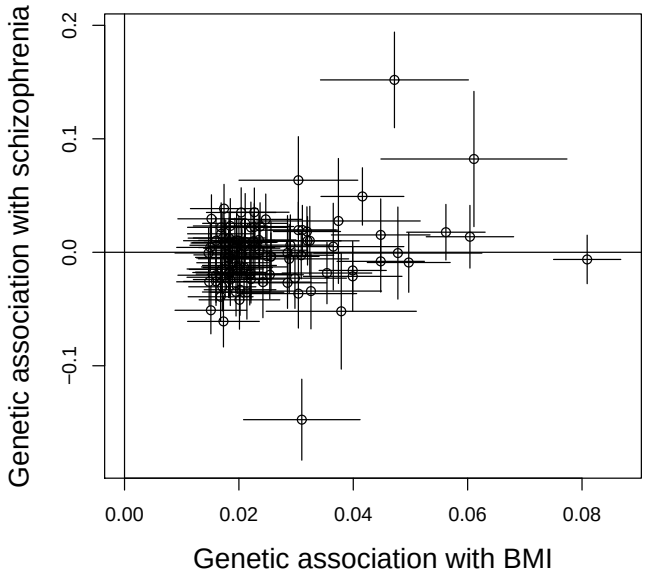


Figure 4: Estimated genetic associations and 95% confidence intervals with body mass index (BMI, standard deviation units) and with schizophrenia risk (log odds ratios) for 97 genetic variants.

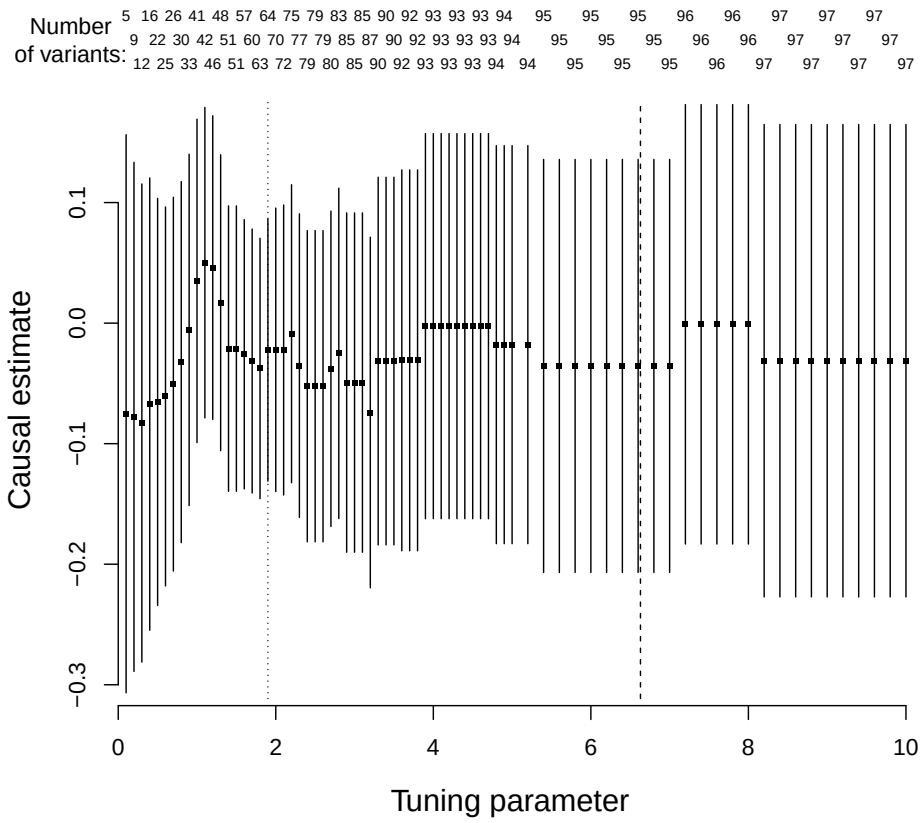


Figure 5: Causal effect estimates (point estimate and 95% confidence interval) for a range of values of the tuning parameter in applied example (BMI on schizophrenia). The number of genetic variants included in each analysis is also displayed. The dotted line at $\lambda = 1.9$ is the value of the tuning parameter chosen by the heterogeneity criterion. The dashed line at $\lambda = 6.63$ is the value chosen by cross-validation.

Web Appendix

A.1 Software code

We provide R code to implement the methods discussed in this paper. The associations of the candidate instruments with the exposure are denoted `betaXG` with standard errors `sebetaXG`. The associations of the candidate instruments with the outcome are denoted `betaYG` with standard errors `sebetaYG`. We assume that the candidate instruments are uncorrelated in their distributions, as is common in applied Mendelian randomization investigations.

Inverse-variance weighted estimate:

The inverse-variance weighted (IVW) estimate can be calculated by weighted linear regression:

```
betaIVW               = summary(lm(betaYG~betaXG-1, weights=sebetaYG^-2))$coef[1]
sebetaIVW.fixed       = summary(lm(betaYG~betaXG-1, weights=sebetaYG^-2))$coef[1,2]/
                           summary(lm(betaYG~betaXG-1, weights=sebetaYG^-2))$sigma
sebetaIVW.random      = summary(lm(betaYG~betaXG-1, weights=sebetaYG^-2))$coef[1,2]/
                      min(summary(lm(betaYG~betaXG-1, weights=sebetaYG^-2))$sigma,1)
```

In the fixed-effect model, we divide the reported standard error by the estimated residual standard error, to fix the residual standard error to take the value 1 [49]. In the multiplicative random-effects model, we divide by the estimated residual standard error in the case of underdispersion (the variability in the genetic associations is less than would be expected by chance alone). But in the case of overdispersion (that is, heterogeneity of causal effect estimates), no correction is made. The point estimate is unaffected by the choice of a fixed- or multiplicative random-effects model.

Alternatively, the inverse-variance weighted estimate can be calculated by meta-analysis, or via a simple formula:

```
# meta-analysis
library(meta)
betaIVW               = metagen(betaYG/betaXG, abs(sebetaYG/betaXG))$TE.fixed
sebetaIVW.fixed       = metagen(betaYG/betaXG, abs(sebetaYG/betaXG))$seTE.fixed
# simple formula
betaIVW               = sum(betaYG*betaXG*sebetaYG^-2)/sum(betaXG^2*sebetaYG^-2)
sebetaIVW.fixed       = 1/sqrt(sum(betaXG^2*sebetaYG^-2))
```

The meta-analysis method can be used to perform an additive random-effects analysis, which makes a different parametric assumption about the heterogeneity between causal estimates compared with the multiplicative random-effect analysis [22]. While the causal estimates from the fixed-effect and multiplicative random-effects analyses are the same, the estimate from the additive random-effects analysis differs.

MR-Egger regression:

The MR-Egger method is equivalent to the IVW method calculated using weighted regression, except that that intercept term is estimated rather than being set to zero. A test as to whether the intercept term is equal to zero is a test of directional pleiotropy. A random-effects model should be used for inference as a fixed-effect model is not justifiable when the candidate instruments are not all valid.

```
# coding of genetic variants
betaYG = betaYG*sign(betaXG); betaXG = abs(betaXG)
# causal estimate
betaEGGER          = summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$coef[2,1]
sebetaEGGER.random = summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$coef[2,2]/
                       min(summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$sigma, 1)
betaEGGER.lower    = betaEGGER-qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random
betaEGGER.upper    = betaEGGER+qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random
p.causal.random    = 2*(1-pt(abs(betaEGGER/sebetaEGGER.random),df=length(betaXG)-2))
# test for directional pleiotropy
interEGGER          = summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$coef[1,1]
seinterEGGER.random = summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$coef[1,2]/
                    min(summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$sigma, 1)
p.dpleio.random     = 2*(1-pt(abs(interEGGER/seinterEGGER.random),df=length(betaXG)-2))
```

In this code, we use a t-distribution with $J - 2$ degrees of freedom for inference. If there is underdispersion, then the t-distribution may be overly conservative, as the t-distribution assumes that the residual standard error is estimated (in case of underdispersion, the residual standard error is set to 1). Hence, if the residual standard error is less than one, either a confidence interval using a residual standard error of 1 and a z-distribution, or else a confidence interval using the estimated residual standard error and a t-distribution may be preferred (the wider of these two intervals should be preferred – both of these will be narrower than the above confidence interval).

```
sigmaEGGER         = summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$sigma
betaEGGER.lower    = ifelse(sigmaEGGER<1, min(betaEGGER-qnorm(0.975)*sebetaEGGER.random,
                   betaEGGER-qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random*sigmaEGGER),
                   betaEGGER-qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random)
betaEGGER.upper    = ifelse(sigmaEGGER<1, max(betaEGGER+qnorm(0.975)*sebetaEGGER.random,
                   betaEGGER+qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random*sigmaEGGER),
                   betaEGGER+qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random)
```

Median-based method:

The median-based method calculates the median (or weighted median) of the causal estimates from each candidate instrument. This code calculates the simple median, weighted median, and penalized weighted median, employing bootstrapping to obtain a standard error that can used to provide a confidence interval.

```
weighted.median <- function(betaIV.in, weights.in) {
  betaIV.order  = betaIV.in[order(betaIV.in)]
  weights.order = weights.in[order(betaIV.in)]
  weights.sum   = cumsum(weights.order)-0.5*weights.order
  weights.sum   = weights.sum/sum(weights.order)
        below   = max(which(weights.sum<0.5))
  weighted.est  = betaIV.order[below] + (betaIV.order[below+1]-betaIV.order[below])*
                  (0.5-weights.sum[below])/(weights.sum[below+1]-weights.sum[below])
```

```
  return(weighted.est) }
 #
weighted.median.boot = function(betaXG.in, betaYG.in, sebetaXG.in, sebetaYG.in, weights.in){
 # the standard error is estimated based on 1000 bootstrap samples
med = NULL
for(i in 1:1000){
 betaXG.boot = rnorm(length(betaXG.in), mean=betaXG.in, sd=sebetaXG.in)
 betaYG.boot = rnorm(length(betaYG.in), mean=betaYG.in, sd=sebetaYG.in)
 betaIV.boot  = betaYG.boot/betaXG.boot
 med[i] = weighted.median(betaIV.boot, weights.in)
 }
return(sd(med)) }
 #
betaIV             = betaYG/betaXG
weights            = rep(1, length(betaXG)) # unweighted median
betaSIMPLEMED      = weighted.median(betaIV, weights)
sebetaSIMPLEMED    = weighted.median.boot(betaXG, betaYG, sebetaXG, sebetaYG, weights)
lowerSIMPLEMED     = betaSIMPLEMED-qnorm(0.975)*sebetaSIMPLEMED
upperSIMPLEMED     = betaSIMPLEMED+qnorm(0.975)*sebetaSIMPLEMED
 #
betaIV             = betaYG/betaXG
weights            = (sebetaYG/betaXG)^-2   # weighted median using inverse-variance weights
betaWEIGHTEDMED    = weighted.median(betaIV, weights)
sebetaWEIGHTEDMED  = weighted.median.boot(betaXG, betaYG, sebetaXG, sebetaYG, weights)
lowerWEIGHTEDMED   = betaWEIGHTEDMED-qnorm(0.975)*sebetaWEIGHTEDMED
upperWEIGHTEDMED   = betaWEIGHTEDMED+qnorm(0.975)*sebetaWEIGHTEDMED
 #
betaIV             = betaYG/betaXG          # penalized weighted median
penalty            = pchisq(weights*(betaIV-betaWEIGHTEDMED)^2, df=1, lower.tail=FALSE)
pen.weights        = (sebetaYG/betaXG)^-2*pmin(1, penalty*20)    # penalized weights
betaPENALIZEDMED   = weighted.median(betaIV, pen.weights)
sebetaPENALIZEDMED = weighted.median.boot(betaXG, betaYG, sebetaXG, sebetaYG, pen.weights)
lowerPENALIZEDMED  = betaPENALIZEDMED-qnorm(0.975)*sebetaPENALIZEDMED
upperPENALIZEDMED  = betaPENALIZEDMED+qnorm(0.975)*sebetaPENALIZEDMED
```

Robust regression:

The IVW and MR-Egger methods can be performed using robust regression (in particular, MM-estimation using Tukey's bisquare objective function) rather than standard linear regression:

```
library(robustbase)
betaIVW.robust            = summary(lmrob(betaYG~betaXG-1, weights=sebetaYG^-2, k.max=500))$coef[1]
sebetaIVW.robust.fixed    = summary(lmrob(betaYG~betaXG-1, weights=sebetaYG^-2, k.max=500))$coef[1,2]/
                            summary(lmrob(betaYG~betaXG-1, weights=sebetaYG^-2, k.max=500))$sigma
sebetaIVW.robust.random   = summary(lmrob(betaYG~betaXG-1, weights=sebetaYG^-2, k.max=500))$coef[1,2]/
                       min(summary(lmrob(betaYG~betaXG-1, weights=sebetaYG^-2, k.max=500))$sigma,1)
betaEGGER.robust          = summary(lmrob(betaYG~betaXG,   weights=sebetaYG^-2, k.max=500))$coef[2]
sebetaEGGER.robust.random = summary(lmrob(betaYG~betaXG,   weights=sebetaYG^-2, k.max=500))$coef[2,2]/
                       min(summary(lmrob(betaYG~betaXG,   weights=sebetaYG^-2, k.max=500))$sigma,1)
```

The `k.max` option sets the maximum number of steps evaluated to find initial parameter values in the S-step of the algorithm.

Penalized weights:

The IVW and MR-Egger methods can be performed using penalized weights:

```
betaIVW            = sum(betaYG*betaXG*sebetaYG^-2)/sum(betaXG^2*sebetaYG^-2)
pweights  = pchisq(betaXG^2/sebetaYG^2*(betaYG/betaXG-betaIVW)^2, df=1, lower.tail=FALSE)
```

```
pweightsE = pchisq(sebetaYG^-2*(betaYG - interEGGER - betaEGGER*betaXG)^2, df=1, lower.tail=FALSE)
rweights  = sebetaYG^-2*pmin(1, pweights*20)
rweightsE = sebetaYG^-2*pmin(1, pweightsE*20)
betaIVW.penal            = summary(lm(betaYG~betaXG-1, weights=rweights))$coef[1]
sebetaIVW.penal.fixed    = summary(lm(betaYG~betaXG-1, weights=rweights))$coef[1,2]/
                           summary(lm(betaYG~betaXG-1, weights=rweights))$sigma
sebetaIVW.penal.random   = summary(lm(betaYG~betaXG-1, weights=rweights))$coef[1,2]/
                       min(summary(lm(betaYG~betaXG-1, weights=rweights))$sigma,1)
betaEGGER.penal          = summary(lm(betaYG~betaXG,   weights=rweightsE))$coef[2]
sebetaEGGER.penal.random = summary(lm(betaYG~betaXG,   weights=rweightsE))$coef[2,2]/
                       min(summary(lm(betaYG~betaXG,   weights=rweightsE))$sigma,1)
```

Penalized weights can also be used in conjunction with robust regression.

L1 penalization:

Several packages are available for running various flavours of L1 penalization methods. We chose the *penalized* package as this gave an option for some of the coefficients in the model to be penalized (the pleiotropy intercept parameters), and others not to be penalized (the causal effect parameter):

```
library(penalized)
betaYGw = betaYG/sebetaYG  # dividing the association estimates by sebetaYG is equivalent
betaXGw = betaXG/sebetaYG  # to weighting by sebetaYG^-2
pleio = diag(rep(1, length(betaXG)))

l1one_which = which(attributes(penalized(betaYGw, pleio, betaXGw, lambda1=1))$penalized==0)
l1one_beta  = lm(betaYG[l1one_which]~betaXG[l1one_which]-1, weights=sebetaYG[l1one_which]^-2)$coef
l1one_se    = summary(lm(betaYG[l1one_which]~betaXG[l1one_which]-1, weights=sebetaYG[l1one_which]^-
         /min(summary(lm(betaYG[l1one_which]~betaXG[l1one_which]-1, weights=sebetaYG[l1one_which]^-

l1two_which = which(attributes(penalized(betaYGw, pleio, betaXGw, lambda1=2))$penalized==0)
l1two_beta  = lm(betaYG[l1two_which]~betaXG[l1two_which]-1, weights=sebetaYG[l1two_which]^-2)$coef
l1two_se    = summary(lm(betaYG[l1two_which]~betaXG[l1two_which]-1, weights=sebetaYG[l1two_which]^-
         /min(summary(lm(betaYG[l1two_which]~betaXG[l1two_which]-1, weights=sebetaYG[l1two_which]^-

l1three_which = which(attributes(penalized(betaYGw, pleio, betaXGw, lambda1=3))$penalized==0)
l1three_beta  = lm(betaYG[l1three_which]~betaXG[l1three_which]-1, weights=sebetaYG[l1three_which]^-
l1three_se    = summary(lm(betaYG[l1three_which]~betaXG[l1three_which]-1, weights=sebetaYG[l1three_
           /min(summary(lm(betaYG[l1three_which]~betaXG[l1three_which]-1, weights=sebetaYG[l1three_
 # fixing lambda to be 1, 2, and 3 in turn

l1grid = c(seq(from=0.1, to=5, by=0.1), seq(from=5.2, to=10, by=0.2))
    # values of lambda for grid search
l1grid_rse = NULL; l1grid_length = NULL; l1grid_beta = NULL; l1grid_se = NULL

for (i in 1:length(l1grid)) {
 l1grid_which = which(attributes(penalized(betaYGw, pleio, betaXGw, lambda1=l1grid[i], trace=FALSE)
 l1grid_rse[i]    = summary(lm(betaYG[l1grid_which]~betaXG[l1grid_which]-1, weights=sebetaYG[l1gric
 l1grid_length[i] = length(l1grid_which)
 l1grid_beta[i]   = lm(betaYG[l1grid_which]~betaXG[l1grid_which]-1, weights=sebetaYG[l1grid_which]
 l1grid_se[i]     = summary(lm(betaYG[l1grid_which]~betaXG[l1grid_which]-1, weights=sebetaYG[l1gric
                min(summary(lm(betaYG[l1grid_which]~betaXG[l1grid_which]-1, weights=sebetaYG[l1gric
 }

l1which_hetero = c(which(l1grid_rse[1:(length(l1grid)-1)]>1&
```

```
                  diff(l1grid_rse)>qchisq(0.95, df=1)/l1grid_length[2:length(l1grid)]), length(l1
    # heterogeneity criterion for choosing lambda
l1hetero_beta  = l1grid_beta[l1which_hetero]
l1hetero_se    = l1grid_se[l1which_hetero]

l1which_min    = which.min(l1grid_beta)
l1min_beta     = l1grid_beta[l1which_min]
l1min_se       = l1grid_se[l1which_min]
    # minimal estimate criterion for choosing lambda

l1xval_lambda = optL1(betaYGw, pleio, betaXGw)$lambda
l1xval_which  = which(attributes(penalized(betaYGw, pleio, betaXGw, lambda1=l1xval_lambda))$penali
l1xval_beta   = summary(lm(alpy[l1xval_which]~alpx[l1xval_which]-1, weights=alpysd[l1xval_which]^-
l1xval_se     = summary(lm(alpy[l1xval_which]~alpx[l1xval_which]-1, weights=alpysd[l1xval_which]^-
            min(summary(lm(alpy[l1xval_which]~alpx[l1xval_which]-1, weights=alpysd[l1xval_which]^-
    # cross-validation criterion for choosing lambda
```

We found that the choice of values of λ for the grid search presented here ($0.1, 0.2, \ldots, 4.9, 5.0, 5.2, 5.4, \ldots, 9.8, 10.0$) worked well in both the simulations and the applied example. However, for different sets of association estimates, a different choice of values may be preferred. Additionally, particularly with large numbers of variants, a more dense choice of values may be preferred to ensure that at must one variant is added to the analysis at each incremental step.

A.2 Choice of penalty function

When there are two candidate instruments, the use of an L1 penalty function is equivalent to minimizing:

$$S_1 = (Y_1 - \theta_{01} - \theta_1 X_1)^2 + (Y_2 - \theta_{01} - \theta_1 X_2)^2 + 2\lambda(|\theta_{01}| + |\theta_{02}|)$$

where Y_j is $\hat{\beta}_{Yj}/\operatorname{se}(\hat{\beta}_{Yj})$ and X_j is $\hat{\beta}_{Xj}/\operatorname{se}(\hat{\beta}_{Yj})$. The factor of two on the penalty function and the change of notation are for simplicity of presentation, and dividing the associations by $\operatorname{se}(\hat{\beta}_{Yj})$ is equivalent to inverse-variance weighting.

Differentiating this expression, we get:

$$\begin{aligned}
\frac{\partial S_1}{\partial \theta_{01}} &= -2(Y_1 - \theta_{01} - \theta_1 X_1) + 2\lambda \operatorname{sign}(\theta_{01}) \\
\frac{\partial S_1}{\partial \theta_{02}} &= -2(Y_2 - \theta_{02} - \theta_1 X_2) + 2\lambda \operatorname{sign}(\theta_{02}) \\
\frac{\partial S_1}{\partial \theta_1} &= -2X_1(Y_1 - \theta_{01} - \theta_1 X_1) - 2X_2(Y_2 - \theta_{02} - \theta X_2)
\end{aligned}$$

As S_1 is not continuous, this function is minimized either at a discontinuity ($\theta_{01} = 0$, $\theta_{02} = 0$), or where the derivatives equal zero. If $\hat{\theta}_{01}$ and $\hat{\theta}_{02}$ both differ from zero, then:

$$\begin{aligned}
\hat{\theta}_{01} &= Y_1 - \hat{\theta}_1 X_1 - \operatorname{sign}(\hat{\theta}_{01})\lambda \\
\hat{\theta}_{02} &= Y_2 - \hat{\theta}_1 X_1 - \operatorname{sign}(\hat{\theta}_{02})\lambda \\
\hat{\theta}_1 &= \frac{-\lambda(X_1 \operatorname{sign}(\hat{\theta}_{01}) + X_2 \operatorname{sign}(\hat{\theta}_{02}))}{2(X_1^2 + X_2^2)}
\end{aligned}$$

When λ is close to zero, $\hat{\theta}_{01}$ and $\hat{\theta}_{02}$ will both differ from zero, whereas when λ is large, $\hat{\theta}_{01}$ and $\hat{\theta}_{02}$ will both equal zero. The upshot is that the value of $\hat{\theta}_1$ depends on the value of λ.

In contrast, if we were to use an L2 penalty function, we would minimize:

$$S_2 = (Y_1 - \theta_{01} - \theta_1 X_1)^2 + (Y_2 - \theta_{01} - \theta_1 X_2)^2 + \lambda(\theta_{01}^2 + \theta_{02}^2)$$

This function is continuous, and so its minimum is where the partial derivatives

equal zero:

$$\begin{aligned}
\frac{\partial S_2}{\partial \theta_{01}} &= -2(Y_1 - \theta_{01} - \theta_1 X_1) + 2\lambda\theta_{01} \\
\frac{\partial S_2}{\partial \theta_{02}} &= -2(Y_2 - \theta_{02} - \theta_1 X_2) + 2\lambda\theta_{02} \\
\frac{\partial S_2}{\partial \theta_1} &= -2X_1(Y_1 - \theta_{01} - \theta_1 X_1) - 2X_2(Y_2 - \theta_{02} - \theta X_2) \\
\hat{\theta}_{01} &= \frac{Y_1 - \hat{\theta}_1 X_1}{1+\lambda} \\
\hat{\theta}_{02} &= \frac{Y_2 - \hat{\theta}_1 X_1}{1+\lambda} \\
\hat{\theta}_1 &= \frac{X_1 Y_1 + X_2 Y_2}{X_1^2 + X_2^2}
\end{aligned}$$

The causal estimate is not a function of λ. Hence, L2 penalization cannot be used either for robust estimation, or for identifying valid instruments (as it does not have a sparsity property; $\hat{\theta}_{01}$ and $\hat{\theta}_{02}$ differ from zero for all finite values of λ).

A.3 Supplementary tables for simulation study

A.3.1 Number of simulations that failed to report a standard error

The numbers of simulations for the robust methods that failed to report a standard error in Scenarios 2 to 4 are provided in Web Table A1. The proportion of simulations was usually less than 1%, and was less than 2.5% in all cases.

Method	Null causal effect			Positive causal effect		
Proportion invalid:	10%	20%	30%	10%	20%	30%
Scenario 2: balanced pleiotropy, InSIDE satisfied						
Robust, no intercept	1	2	0	0	1	0
Robust, intercept	5	20	24	12	18	24
Penalized robust, no intercept	4	5	18	0	10	12
Penalized robust, intercept	15	42	97	10	23	80
Scenario 3: directional pleiotropy, InSIDE satisfied						
Robust, no intercept	1	0	0	0	1	1
Robust, intercept	2	18	15	5	10	11
Penalized robust, no intercept	4	5	8	1	2	3
Penalized robust, intercept	2	30	44	15	19	31
Scenario 4: directional pleiotropy, InSIDE not satisfied						
Robust, no intercept	2	15	34	3	15	26
Robust, intercept	131	245	244	147	233	227
Penalized robust, no intercept	2	12	44	2	9	41
Penalized robust, intercept	24	69	102	31	51	92

Web Table A1: Number of the 10 000 simulations that failed to report a standard error using the robust regression method in each of the simulation settings.

A.3.2 One-sample setting

The simulation study from the main body of the paper was repeated, except in a one-sample setting in which associations of the candidate instruments with the exposure and with the outcome were obtained in the same sample of 20 000 individuals for the methods using non-penalized weights. Results are displayed in Web Table A2 (Scenario 1) and Web Table A3 (Scenarios 2 to 4).

A.3.3 Fewer candidate instruments

The simulation was also repeated in a two-sample setting with only 10 candidate instruments, to observe whether the robust methods were able to operate well with fewer instruments to detect violations of the instrumental variables assumptions. Results for Scenarios 2 to 4 are presented in Web Table A4.

Method	Scenario 1 Mean	SD	Mean SE	Power	NA [1]
Null causal effect: $\theta = 0$					
Standard, no intercept [2]	0.024	0.044	0.047	6.8	-
Standard, intercept [3]	0.173	0.123	0.131	27.2	-
Robust, no intercept	0.023	0.044	0.049	7.8	0
Robust, intercept	0.174	0.126	0.136	29.0	4
Penalized standard, no intercept	0.024	0.045	0.046	8.5	-
Penalized standard, intercept	0.174	0.126	0.129	28.5	-
Penalized robust, no intercept	0.024	0.046	0.047	9.4	0
Penalized robust, intercept	0.174	0.127	0.131	31.0	3
Simple median	0.000	0.060	0.070	2.0	-
Weighted median	0.038	0.054	0.063	5.5	-
Penalized weighted median	0.038	0.057	0.063	6.5	-
Positive causal effect: $\theta = +0.1$					
Standard, no intercept [2]	0.123	0.044	0.048	73.6	-
Standard, intercept [3]	0.271	0.121	0.136	53.1	-
Robust, no intercept	0.122	0.045	0.050	69.3	0
Robust, intercept	0.271	0.125	0.143	52.4	3
Penalized standard, no intercept	0.122	0.045	0.048	74.1	-
Penalized standard, intercept	0.271	0.123	0.135	53.8	-
Penalized robust, no intercept	0.122	0.046	0.049	71.6	1
Penalized robust, intercept	0.271	0.126	0.139	53.9	2
Simple median	0.099	0.059	0.073	25.4	-
Weighted median	0.136	0.054	0.067	54.6	-
Penalized weighted median	0.136	0.057	0.067	54.5	-

Web Table A2: Mean, standard deviation (SD), mean standard error (mean SE) of estimates, and empirical power (%) from weighted linear regression models (weights are penalized where indicated) using standard and robust regression, without and with an intercept term, and median-based methods for Scenario 1 in one-sample setting (associations with exposure and with outcome are estimated in the same individuals).

[1] Number of the 10 000 simulations that failed to report a standard error.
[2] This is the standard inverse-variance weighted (IVW) method.
[3] This is the MR-Egger method.

Method	Scenario 2			Scenario 3			Scenario 4		
	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power
Null causal effect: $\theta = 0$									
Proportion of invalid instrumental variables: 0.1									
Standard, no intercept	0.023	0.069	7.3	0.090	0.066	24.5	0.081	0.072	27.1
Standard, intercept	0.174	0.197	19.4	0.175	0.190	19.2	0.276	0.228	50.7
Robust, no intercept	0.023	0.052	8.3	0.045	0.053	13.2	0.039	0.057	10.4
Robust, intercept	0.174	0.148	26.9	0.175	0.148	27.1	0.263	0.187	36.9
Simple median	-0.001	0.066	2.9	0.028	0.064	4.5	0.012	0.065	3.2
Weighted median	0.037	0.061	6.8	0.061	0.060	12.6	0.074	0.074	20.0
Proportion of invalid instrumental variables: 0.2									
Standard, no intercept	0.022	0.087	6.8	0.158	0.082	50.4	0.133	0.084	44.3
Standard, intercept	0.172	0.248	13.9	0.178	0.232	14.9	0.354	0.243	58.8
Robust, no intercept	0.023	0.063	7.7	0.084	0.072	19.4	0.072	0.078	15.5
Robust, intercept	0.172	0.187	23.0	0.175	0.186	22.6	0.382	0.225	51.9
Simple median	-0.001	0.073	3.5	0.064	0.073	11.2	0.026	0.071	4.7
Weighted median	0.037	0.067	7.8	0.089	0.069	22.5	0.121	0.096	39.0
Proportion of invalid instrumental variables: 0.3									
Standard, no intercept	0.024	0.102	6.8	0.226	0.091	74.7	0.181	0.088	58.8
Standard, intercept	0.171	0.286	11.0	0.178	0.259	12.8	0.404	0.243	61.8
Robust, no intercept	0.022	0.080	6.7	0.143	0.096	30.3	0.120	0.098	26.8
Robust, intercept	0.172	0.234	17.5	0.176	0.232	17.6	0.485	0.239	65.5
Simple median	-0.001	0.082	4.7	0.109	0.085	24.7	0.043	0.080	8.0
Weighted median	0.036	0.077	9.2	0.127	0.084	38.4	0.176	0.117	58.7
Positive causal effect: $\theta = +0.1$									
Proportion of invalid instrumental variables: 0.1									
Standard, no intercept	0.123	0.069	49.4	0.190	0.066	87.2	0.181	0.072	79.3
Standard, intercept	0.274	0.197	35.7	0.275	0.190	36.4	0.376	0.228	65.1
Robust, no intercept	0.123	0.052	62.8	0.145	0.053	74.8	0.139	0.057	67.3
Robust, intercept	0.274	0.148	49.2	0.275	0.148	49.2	0.363	0.187	55.7
Simple median	0.099	0.066	24.6	0.128	0.064	38.9	0.112	0.065	30.2
Weighted median	0.137	0.061	52.0	0.161	0.060	66.0	0.174	0.074	70.1
Proportion of invalid instrumental variables: 0.2									
Standard, no intercept	0.123	0.087	34.9	0.258	0.082	95.2	0.233	0.084	84.8
Standard, intercept	0.272	0.248	25.2	0.278	0.232	27.2	0.454	0.243	69.9
Robust, no intercept	0.123	0.063	52.4	0.184	0.072	76.3	0.172	0.078	64.8
Robust, intercept	0.273	0.187	40.7	0.275	0.186	40.5	0.482	0.225	64.3
Simple median	0.099	0.073	23.9	0.164	0.073	54.6	0.126	0.071	35.1
Weighted median	0.137	0.067	49.8	0.189	0.069	76.4	0.221	0.096	81.4
Proportion of invalid instrumental variables: 0.3									
Standard, no intercept	0.124	0.102	26.8	0.326	0.091	98.1	0.281	0.088	89.3
Standard, intercept	0.271	0.286	19.0	0.278	0.259	21.4	0.504	0.243	73.3
Robust, no intercept	0.122	0.080	40.2	0.243	0.096	76.8	0.220	0.098	66.7
Robust, intercept	0.272	0.234	30.6	0.276	0.232	31.0	0.585	0.239	74.3
Simple median	0.099	0.082	23.2	0.209	0.085	69.5	0.143	0.080	40.8
Weighted median	0.136	0.077	46.6	0.227	0.084	84.6	0.276	0.117	89.1

Web Table A3: Mean, standard deviation (SD), mean standard error (mean SE) of estimates, and empirical power (%) from weighted linear regression models (weights are not penalized) using standard and robust regression, without and with an intercept term, and simple and weighted median methods for Scenarios 2, 3, and 4 in one-sample setting (associations with exposure and with outcome are estimated in the same individuals).

Method	Scenario 2			Scenario 3			Scenario 4		
	Mean	SD	Power	Mean	SD	Power	Mean	SD	Power
Null causal effect: $\theta = 0$									
Proportion of invalid instrumental variables: 0.1									
Standard, no intercept	-0.001	0.113	5.5	0.068	0.109	7.2	0.053	0.112	11.4
Standard, intercept	-0.003	0.353	6.2	0.001	0.343	6.3	0.123	0.366	20.0
Robust, no intercept	0.000	0.090	7.0	0.030	0.093	7.4	0.024	0.098	8.4
Robust, intercept	0.001	0.321	12.5	0.002	0.315	12.3	0.092	0.348	19.7
Simple median	-0.001	0.105	3.2	0.033	0.106	4.0	0.013	0.103	3.0
Weighted median	0.000	0.098	4.2	0.028	0.100	4.9	0.044	0.124	11.0
Proportion of invalid instrumental variables: 0.2									
Standard, no intercept	0.001	0.139	5.8	0.136	0.133	13.9	0.101	0.133	18.4
Standard, intercept	0.001	0.432	7.1	0.002	0.412	7.5	0.213	0.418	30.8
Robust, no intercept	0.002	0.111	7.2	0.076	0.126	9.9	0.058	0.125	12.6
Robust, intercept	0.002	0.398	14.3	0.001	0.387	14.2	0.185	0.418	29.5
Simple median	0.000	0.117	4.0	0.075	0.129	8.2	0.028	0.117	5.0
Weighted median	0.001	0.113	5.8	0.063	0.125	9.9	0.092	0.152	21.8
Proportion of invalid instrumental variables: 0.3									
Standard, no intercept	0.001	0.166	7.0	0.205	0.151	24.7	0.148	0.145	25.3
Standard, intercept	0.011	0.510	7.8	0.008	0.461	7.8	0.287	0.426	35.7
Robust, no intercept	0.000	0.140	7.8	0.140	0.160	15.9	0.102	0.148	20.4
Robust, intercept	0.002	0.494	13.4	0.003	0.455	14.7	0.271	0.453	36.7
Simple median	0.001	0.140	6.3	0.134	0.164	17.6	0.052	0.136	8.6
Weighted median	0.001	0.136	8.6	0.110	0.156	17.7	0.144	0.173	34.5
Positive causal effect: $\theta = +0.1$									
Proportion of invalid instrumental variables: 0.1									
Standard, no intercept	0.095	0.116	18.1	0.164	0.112	30.3	0.149	0.116	32.1
Standard, intercept	0.063	0.361	6.8	0.067	0.352	6.7	0.196	0.376	22.8
Robust, no intercept	0.096	0.095	20.8	0.128	0.098	26.1	0.122	0.103	25.2
Robust, intercept	0.066	0.337	13.3	0.068	0.328	13.7	0.160	0.363	21.8
Simple median	0.101	0.111	11.8	0.136	0.113	17.9	0.115	0.109	14.2
Weighted median	0.093	0.104	13.0	0.122	0.106	18.8	0.140	0.130	25.3
Proportion of invalid instrumental variables: 0.2									
Standard, no intercept	0.097	0.141	15.4	0.232	0.136	40.9	0.197	0.136	39.7
Standard, intercept	0.067	0.439	7.4	0.068	0.420	8.1	0.292	0.427	34.5
Robust, no intercept	0.097	0.116	18.7	0.175	0.130	30.3	0.157	0.129	30.1
Robust, intercept	0.068	0.411	14.9	0.067	0.401	15.1	0.259	0.432	31.8
Simple median	0.101	0.123	12.1	0.180	0.136	26.7	0.130	0.124	17.3
Weighted median	0.094	0.118	14.0	0.159	0.130	26.7	0.188	0.158	37.7
Proportion of invalid instrumental variables: 0.3									
Standard, no intercept	0.097	0.167	14.4	0.301	0.154	53.0	0.244	0.147	46.9
Standard, intercept	0.077	0.516	8.3	0.074	0.467	8.0	0.371	0.434	39.9
Robust, no intercept	0.096	0.144	16.9	0.240	0.162	38.4	0.202	0.151	37.0
Robust, intercept	0.068	0.505	14.4	0.068	0.464	15.5	0.349	0.465	39.4
Simple median	0.102	0.146	14.1	0.240	0.171	39.9	0.155	0.142	23.4
Weighted median	0.094	0.142	16.4	0.207	0.160	38.1	0.242	0.178	51.2

Web Table A4: Mean, standard deviation (SD), mean standard error (mean SE) of estimates, and empirical power (%) from weighted linear regression models (weights are not penalized) using standard and robust regression, without and with an intercept term, and simple and weighted median methods for Scenarios 2, 3, and 4 in two-sample setting with only 10 candidate instruments (25 candidate instruments are used in all other simulations).

Appendix B

Paper 1: Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates

Published paper based on the work in Chapter 3.

PLOS ONE

RESEARCH ARTICLE

Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates

Jessica M. B. Rees[1,2], **Angela M. Wood**[1], **Frank Dudbridge**[3], **Stephen Burgess**[1,4]*

1 Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, United Kingdom, **2** Edinburgh Clinical Trials Unit, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH16 4UX, United Kingdom, **3** Department of Health Sciences, University of Leicester, Leicester, LE1 7RH, United Kingdom, **4** MRC Biostatistics Unit, University of Cambridge, Cambridge, CB2 0SR, United Kingdom

* sb452@medschl.cam.ac.uk

Citation: Rees JMB, Wood AM, Dudbridge F, Burgess S (2019) Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates. PLoS ONE 14(9): e0222362. https://doi.org/10.1371/journal.pone.0222362

Editor: N. Charlotte Onland-Moret, Universitair Medisch Centrum Utrecht, NETHERLANDS

Received: April 6, 2019

Accepted: August 27, 2019

Published: September 23, 2019

Data Availability Statement: All data can be found in the PhenoScanner database (http://www.phenoscanner.medschl.cam.ac.uk/). R code for performing the approaches outlined in the paper, and extracting genetic association estimates are found in S1 Appendix.

Funding: Stephen Burgess is supported by Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 204623/Z/16/Z).

Competing interests: I have read the journal's policy and the authors of this manuscript have the

Abstract

Methods have been developed for Mendelian randomization that can obtain consistent causal estimates under weaker assumptions than the standard instrumental variable assumptions. The median-based estimator and MR-Egger are examples of such methods. However, these methods can be sensitive to genetic variants with heterogeneous causal estimates. Such heterogeneity may arise from over-dispersion in the causal estimates, or specific variants with outlying causal estimates. In this paper, we develop three extensions to robust methods for Mendelian randomization with summarized data: 1) robust regression (MM-estimation); 2) penalized weights; and 3) Lasso penalization. Methods using these approaches are considered in two applied examples: one where there is evidence of over-dispersion in the causal estimates (the causal effect of body mass index on schizophrenia risk), and the other containing outliers (the causal effect of low-density lipoprotein cholesterol on Alzheimer's disease risk). Through an extensive simulation study, we demonstrate that robust regression applied to the inverse-variance weighted method with penalized weights is a worthwhile additional sensitivity analysis for Mendelian randomization to provide robustness to variants with outlying causal estimates. The results from the applied examples and simulation study highlight the importance of using methods that make different assumptions to assess the robustness of findings from Mendelian randomization investigations with multiple genetic variants.

Introduction

Mendelian randomization uses genetic variants as instrumental variables to estimate the causal effect of a risk factor on an outcome using observational data [1, 2]. The genetic variants must satisfy the following criteria (illustrated in Fig 1) to be a valid instrumental variable (IV):

- IV1: the variant is associated with the exposure X,

following competing interests: Stephen Burgess is a paid statistical reviewer for PLOS Medicine. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

- IV2: the variant is independent of all confounders U of the exposure-outcome association, and
- IV3: the variant is independent of the outcome Y conditional on the exposure X and confounders U [3].

A recent development in Mendelian randomization is the availability of summarized data: this consists of the associations (beta-coefficients and standard errors) of genetic variants with the risk factor and with the outcome from regressing each variant in turn [4]. Summarized data can be used to calculate an estimate of the causal effect of the risk factor on the outcome for each genetic variant. The inverse-variance weighted (IVW) method [5] combines these estimates to provide an overall estimate of the causal effect using summarized data from all the genetic variants. If the genetic variants are uncorrelated, the IVW estimate is asymptotically equal to the estimate from the two-stage least squares method commonly used with individual-level data [6].

The inclusion of a variant in a Mendelian randomization analysis that violates either the IV2 or the IV3 assumption may lead to biased causal estimates [7]. Robust methods have therefore been developed to estimate consistent causal effects under weaker assumptions when there are multiple genetic variants. These methods include a median-based method [8] and MR-Egger [9]. Genetic variants that violate the IV assumptions are likely to have heterogeneous causal estimates. We here consider heterogeneity in two settings: firstly, when there is more variance between the variant-specific causal estimates than expected by chance, but the burden of heterogeneity is shared across several genetic variants (over-dispersion); and secondly, when specific variants have outlying causal estimates, and they alone are responsible for driving the observed heterogeneity.

Several robust methods have been proposed that try to identify and remove genetic variants with heterogeneous causal estimates that are suspected to be invalid instruments. These include the MR-PRESSO [10], global and individual tests for direct effects (GLIDE) [11], and generalized summary Mendelian randomization (GSMR) [12] methods. Cochran's Q-statistic has been used in Mendelian randomization to downweight [8] or exclude genetic variants with heterogeneous causal estimates [13]. The Q-statistic is based on the first order weights of the IVW model which assumes that there is no measurement error (NOME) in the genetic

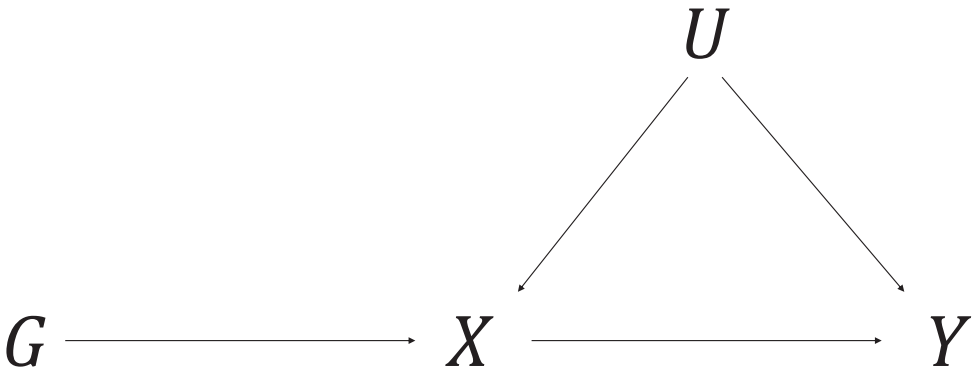


Fig 1. Causal directed acyclic graph illustrating the instrumental variable assumptions for the instrumental variable *G*, exposure *X*, outcome *Y*, and the set of variables (*U*) that confound the association between *X* and *Y*.

https://doi.org/10.1371/journal.pone.0222362.g001

associations with the risk factor [14]. If this assumption is invalid, then the type I error rate of the Q-statistic will be inflated. Bowden *et al.* [14] have accounted for possible violations in the NOME assumption by using adapted second order weights to calculate the Q-statistic.

We here propose three further ways of downweighting or excluding variants with heterogeneous causal estimates that could be considered as part of a sensitivity analysis in a Mendelian randomization study. The first two of these extensions can be used as modifications to either the IVW or the MR-Egger method. These extensions have been influenced by the literature on robust statistics [15], and recent developments in robust methods for Mendelian randomization.

First, we outline the parametric assumptions made throughout the paper and discuss the estimation of the causal effect in a Mendelian randomization study. We then introduce three robust approaches: robust regression (MM-estimation), penalized weights, and Lasso penalization. We apply these approaches to published data on body mass index (BMI) and schizophrenia risk, and on low-density lipoprotein cholesterol (LDL-C) and Alzheimer's disease (AD) risk. Next, we perform a simulation study under realistic settings to compare bias and coverage properties of the robust methods when some of the genetic variants are invalid IVs. Finally, we discuss the results of the paper and its implications to applied Mendelian randomization research. Software code for implementing all of the methods used in this paper, including extracting the genetic association estimates for the applied examples, is provided in S1 Appendix. The methods (excluding Lasso penalization) can also be applied using the R package *MendelianRandomization* [16].

Methods

Parametric assumptions

Throughout the paper, we assume linearity and no effect modification of the causal effect θ of the risk factor on the outcome, and the associations of the genetic variants G_j ($j = 1, \ldots, J$) with the risk factor and with the outcome. These assumptions are not necessary to estimate a causal effect, but they ensure that all valid IVs estimate the same causal parameter. Under these assumptions, the association β_{Y_j} between the variant G_j and the outcome can be decomposed into an indirect effect via the risk factor and a direct (pleiotropic) effect α_j (illustrated in Fig 2):

$$\beta_{Y_j} = \alpha_j + \theta\beta_{X_j}\,. \tag{1}$$

We also assume that the outcome is a continuous variable. If the outcome is binary, then the methods can be applied to the log odds ratios obtained from logistic regression of each genetic variant on the outcome. The linearity assumption must now hold for the logit-transformed probability of the outcome. Difficulties with interpreting the causal estimate of an odds ratio with a binary outcome and a logistic-linear model have been widely discussed [17], with evidence to suggest that the causal estimates tend to be unbiased under the null [18].

Estimating the causal effect

The causal effect θ can be estimated using the genetic associations with the risk factor ($\hat{\beta}_{X_j}$) and with the outcome ($\hat{\beta}_{Y_j}$). The ratio estimate of the causal effect for variant j is given by:

$$\hat{\theta}_j = \frac{\hat{\beta}_{Y_j}}{\hat{\beta}_{X_j}}\,. \tag{2}$$

The J ratio estimates can be combined to provide an overall causal estimate by fitting weighted

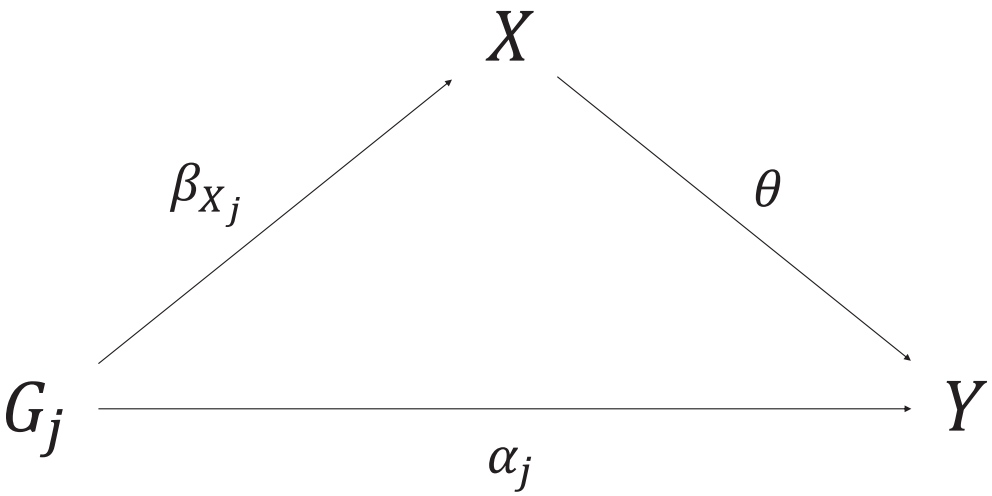


Fig 2. Decomposition of the association between the genetic variant G_j and the outcome Y into the indirect effect via the risk factor X and direct (pleiotropic) effect α_j.

https://doi.org/10.1371/journal.pone.0222362.g002

linear regression of the associations of the variants with the outcome on the associations of the variants with the exposure, with the intercept set to zero and $\text{se}(\hat{\beta}_{Y_j})^{-2}$ as weights:

$$\hat{\beta}_{Y_j} = \theta\,\hat{\beta}_{X_j} + \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0, \psi^2\,\text{se}(\hat{\beta}_{Y_j})^2)\,. \tag{3}$$

The estimate obtained from Eq (3) is equivalent to the estimate from the IVW method [5]. Under a fixed-effects model, we set the residual standard error (ψ) to be equal to one by dividing the standard error of the causal estimate by the estimated residual standard error. To account for heterogeneity (overdispersion) in the causal estimates, the residual standard error can be greater than one under a random-effects model. The causal estimate from the fixed and multiplicative random-effects models will be the same, but the standard error of the causal effect will be larger from the multiplicative random-effects model if there is heterogeneity between the causal estimates.

A genetic variant is pleiotropic if it has a direct effect on the outcome that is not via the risk factor ($\alpha_j \neq 0$). The IVW method under a fixed or multiplicative random-effects model will produce a consistent causal estimate when there is no pleiotropy ($\alpha_j = 0$ for all variants), or when the average pleiotropic effect is zero (referred to as balanced pleiotropy) and the pleiotropic effects are distributed independently of the associations of the genetic variants with the risk factor (known as the InSIDE assumption—Instrument strength independent of the Direct Effect) [9, 19]. If an intercept term in Eq (3) is estimated, then this is the MR-Egger method, and the causal estimate will be consistent in the presence of directional pleiotropy (the average pleiotropic effect differs from zero) if the InSIDE assumption is satisfied [9]:

$$\hat{\beta}_{Y_j} = \theta_0 + \theta_1\,\hat{\beta}_{X_j} + \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0, \psi_E^2\,\text{se}(\hat{\beta}_{Y_j})^2)\,. \tag{4}$$

If the genetic variants are all valid IVs, then the ratio estimates for each variant should be similar. If the $\hat{\beta}_{Y_j}$ estimates were plotted against the $\hat{\beta}_{X_j}$ estimates, a pleiotropic variant may

appear as an outlier relative to the valid IVs as the direct effect of the pleiotropic variant will result in the vertical displacement of $\hat{\beta}_{Y_j}$ from the causal effect (Eq (1)). Robust methods that downweight the contribution of variants with heterogeneous ratio estimates should reduce the impact that variants with outlying or over-dispersed estimates have on the causal estimate. For example, the simple median estimator is the median of the J ratio estimates θ_j ($j = 1, \ldots, J$), and will produce consistent causal estimates if at least 50% of the genetic variants are valid IVs [8].

Typically, applied Mendelian randomization analyses will use one variant from each gene region. Under Mendel's second law, these variants should be independently distributed due to their physical separation. The methods discussed in this paper will therefore assume that the variants are uncorrelated.

Robust regression (MM–estimation). The breakdown point is a measure of the robustness of an estimator to contaminations (such as outliers) in the dataset [15]. Ordinary least squares (OLS) has a breakdown point of 0% as all of the observations have equal weight and just one outlying observation can heavily influence the estimator, resulting in an arbitrarily large or small estimate. Robust regression methods, such as MM-estimation, have been proposed where the breakdown point is greater than 0% [15].

In this paper, we use an MM-estimation approach proposed by Koller and Stahel [20] as it retains the high asymptotic efficiency of the M-estimator ('maximum likelihood type'), whilst utilising the S-estimator ('scale-type estimate') to provide robustness against outliers and leverage points. Under this method, a S-estimate is fitted to minimize the M-estimate of scale, which has the desired high breakdown point but may lack efficiency. The estimates for the scale and regression parameters obtained in this stage are then used to fit an M-estimator with high efficiency, where the scale estimate is held constant to retain the high-breakdown point [20].

Additional robustness in MM-estimation may be achieved by using Tukey's bisquare objective function in the estimation procedure with its weighting function:

$$w(r_j) = \begin{cases} \left[1 - \left(\frac{r_j}{c}\right)^2\right]^2 & \text{if } |r_j| < c \\ 0 & \text{if } |r_j| \geq c \end{cases},$$

where r_j are the standardized residuals, and $w(r_j)$ are used in the objective function of the iteratively reweighted least squares algorithm to obtain the MM-estimates. The recommended values for the tuning parameter c maintain a high breakdown point in the S-estimation step ($c = 1.548$) and provide efficiency in the M-estimation step ($c = 4.685$). In MM-estimation with Tukey's bisquare objective function, the weight of an observation decreases as r_j tends away from zero, and when $|r_j| \geq c$ the observation will have zero weight.

Throughout the paper, we will refer to this approach as robust regression. It is the default implementation of robust regression for the `lmrob` command in the R package *robustbase* [21]. Since the `lmrob` command allows the user to specify a vector of weights to be used in conjunction with Tukey's weighting function, robust regression can be used instead of 'ordinary regression' (weighted least squares) for the IVW and MR-Egger methods.

Penalized weights. We assume that the NOME assumption is satisfied, and propose an approach for downweighting genetic variants with heterogeneous ratio estimates in the IVW model using Cochran's Q statistic:

$$Q = \sum_j Q_j = \sum_j \text{se}\,(\hat{\beta}_{Y_j})^{-2}(\hat{\beta}_{Y_j} - \hat{\theta}\hat{\beta}_{X_j})^2\,, \tag{5}$$

which has an approximate χ^2_{J-1} distribution under the null hypothesis that all J genetic variants satisfy the IV assumptions, with the J components Q_j ($j = 1, \ldots, J$) having approximate χ^2_1 distributions [13]. Since penalized weights would normally be considered when pleiotropy is suspected, the simple (unweighted) median estimate is used for the value of $\hat{\theta}$ in Eq (5) rather than the IVW estimate.

To ensure that the weights $(\mathrm{se}\,(\hat{\beta}_{Y_j})^{-2})$ for the majority of the variants remain the same, we use a penalization for the IVW method based on the one-sided upper tail probability (denoted q_j) of Q_j on a χ^2_1 distribution by multiplying the weights by $\min(1, 100q_j)$. A similar down-weighting factor, $\min(1, 20q_j)$, was used for the penalized–median estimator in the paper by Bowden *et al.* [8]. Initially we used $\min(1, 20q_j)$ but found that too many variants were being penalized, resulting in over-precise estimates that had poor coverage of the true causal effect. By multiplying the weights by $\min(1, 100q_j)$, the outlying variants should be severely penalized, without downweighting too many genetic variants that are valid IVs.

For the MR-Egger method, we consider the modified Q' statistic [22]:

$$Q' = \sum_j Q'_j = \sum_j \mathrm{se}\,(\hat{\beta}_{Y_j})^{-2}(\hat{\beta}_{Y_j} - \hat{\theta}_0 - \hat{\theta}_1\hat{\beta}_{X_j})^2\,, \tag{6}$$

where $\hat{\theta}_0$ and $\hat{\theta}_1$ are taken from the MR-Egger model. If the MR-Egger model is correct, the Q' statistic in Eq (6) should follow an approximate χ^2_{J-2} distribution [23]. The penalized weights described in this Section can also be applied to robust regression for the IVW and MR-Egger methods, subsequently referred to as the robust and penalized approach (or robust regression with penalized weights).

Lasso penalization. The application of Lasso regression in IV analyses has already been considered in the literature [24–26]. The penalty term in Lasso regression shrinks the regression coefficients towards zero, and forces some coefficients to be zero [27]. The sparsity property (shrinking some coefficients to zero) of Lasso regression has been used to identify and remove invalid IVs. The IV methods that use Lasso regression have only been considered with respect to individual level data.

We take the 'post-lasso' method proposed by Windmeijer *et al.* [25] for individual level data and adapt this method to be used with summary level data. First, we consider the objective function for the MR-Egger model that is minimized when fitting the regression model of Eq (4):

$$\sum_j \mathrm{se}\,(\hat{\beta}_{Y_j})^{-2}(\hat{\beta}_{Y_j} - \theta_0 - \theta_1\hat{\beta}_{X_j})^2\,.$$

To better model the pleiotropic effects α_j in Eq (1), we propose replacing θ_0 with a separate intercept coefficient for each genetic variant θ_{0_j}, and adding a Lasso-penalty term for the θ_{0_j} parameters:

$$\sum_j \mathrm{se}\,(\hat{\beta}_{Y_j})^{-2}(\hat{\beta}_{Y_j} - \theta_{0_j} - \theta_1\hat{\beta}_{X_j})^2 + \lambda\sum_j|\theta_{0_j}|\,. \tag{7}$$

If θ_{0_j} shrinks to zero in Eq (7), the genetic variant is treated as a valid IV. We take the genetic variants with a zero intercept term θ_{0_j}, and perform the IVW method using these variants only to estimate the causal effect θ. The degree of shrinkage in Eq (7) is determined by the value of the tuning parameter λ. If $\lambda = \infty$, then all of the genetic variants are assumed to be valid instruments as θ_{0_j} is forced to be zero for all J variants, and the IVW method is

performed using the full set of genetic variants. If $\lambda = 0$, then all of the variants can be pleiotropic, and the parameters in Eq (7) are not identified.

To determine the value of λ, two rules were considered: 1) a heterogeneity stopping rule; and 2) a cross-validation rule. The heterogeneity stopping rule is influenced by the method used by Windmeijer *et al*. [25] and Cochran's Q statistic. For the heterogeneity stopping rule, we fit the Lasso penalization model (Eq (7)) over a range of values for λ, starting with a value close to zero, and then increasing λ in small increments. We stop at $\lambda = \lambda_n$ when the residual standard error from the IVW model, based on the variants determined to be valid from $\lambda = \lambda_{n+1}$, is greater than 1, and the increase in the residual standard error from λ_n to λ_{n+1} is greater than $\chi^2_1(0.95)/J_{inc}$, where $\chi^2_1(0.95)$ is the upper 95th percentile of a chi-squared distribution on 1 degree of freedom, and J_{inc} is the number of genetic variants included in the IVW model when $\lambda = \lambda_{n+1}$.

As an alternative to the heterogeneity stopping rule, we use the `optL1` command in the R package *penalized* [28]. `optL1` compares the predictive ability of the Lasso regression model for different values of λ through leave-one-out cross-validation. The optimal value of λ is then determined by maximizing the cross-validated likelihood function.

Summary

In this Section, we have introduced three robust approaches that can be used in a Mendelian randomization study as part of the sensitivity analysis. The approaches use summary level data that either downweight or remove genetic variants that have heterogeneous causal ratio estimates. In the next Section, we apply these approaches to published summary data to investigate the causal effect of body mass index on schizophrenia risk, and the causal effect of low-density lipoprotein cholesterol on Alzheimer's disease risk.

Applied examples

To illustrate the performance of the proposed extensions, we considered two applied examples: one where there was evidence of over-dispersion in the ratio estimates (the causal effect of BMI on schizophrenia risk); and another that contained outliers (the causal effect of LDL-C on AD risk). Using summary data (beta–coefficients and standard errors) from PhenoScanner [29], we considered the IVW method with: 1) the full set of genetic variants; 2) robust regression; 3) penalized weights; and 4) robust regression and penalized weights. Lasso penalization with the heterogeneity stopping and cross-validation rules, the simple median, the weighted median, and the MR-Egger methods were also considered. Under the heterogeneity stopping rule, the Lasso penalization model was applied to $\lambda = 0.1, 0.2, \ldots, 4.9, 5.0, 5.2, 5.4, \ldots, 9.8, 10.0$. Multiplicative random-effects models were used in all analyses.

Causal effect of body mass index on schizophrenia risk

Although individuals with schizophrenia tend to be overweight [30], it is generally believed that this is due to the effect of anti-psychotic medication on body composition (reverse causation) rather than any causal effect of BMI on schizophrenia risk [31]. For this Mendelian randomization analysis, we used the 97 genetic variants reported by the Genetic Investigation of Anthropometric Traits (GIANT) consortium that were associated with BMI in 339,224 European-descent individuals at a genome-wide level of significance (p-value $< 5 \times 10^{-8}$) [32]. Variants were clumped at a correlation threshold of $r^2 > 0.1$, and all 97 variants are separated by at least 500 kilobases. The genetic associations with schizophrenia were obtained from the Psychiatric Genomics Consortium (PGC) based on 35,476 cases and 46,839 controls mostly of European descent [33]. The summarized data used in this paper were recently applied in a

Mendelian randomization study investigating the causal effect of BMI on psychiatric disorders, including schizophrenia risk [34].

Causal effect of low-density lipoprotein cholesterol on Alzheimer's disease risk

Epidemiological studies have provided evidence of an association between LDL-C and increased risk of AD [35, 36]. However, there is also evidence to suggest that patients with AD have altered lipid metabolism (reverse causation) [37]. In this Mendelian randomization analysis, we used the 75 genetic variants previously demonstrated to be associated with LDL-C at a genome-wide level of significance by the Global Lipids Genetics Consortium (GLGC) [38]. The point estimates for the genetic associations with LDL-C were taken from the linear regression in up to 188,578 participants from GLGC [39]. The majority of variants are separated by at least 1 megabase. A second variant from a gene region was only selected if it was independently associated with LDL-C and in low linkage disequilibrium with the lead variant ($r^2 < 0.05$). A recent Mendelian randomization study used summarized data from GLGC to investigate the causal association between low LDL-C levels and AD risk using data on 380 variants. Our analysis is based on a smaller set of genetic variants compared to Benn *et al.* [40] as we excluded variants that were associated with LDL-C and high-density lipoprotein and/or triglycerides. The genetic associations with AD were obtained from the International Genomics of Alzheimer's Project (IGAP) based on 17,008 cases and 37,154 controls of European-descent [41].

Results

The estimated genetic associations with 95% confidence intervals for the two examples are displayed in Fig 3. The plots demonstrate the overdispersion in the ratio estimates for BMI and

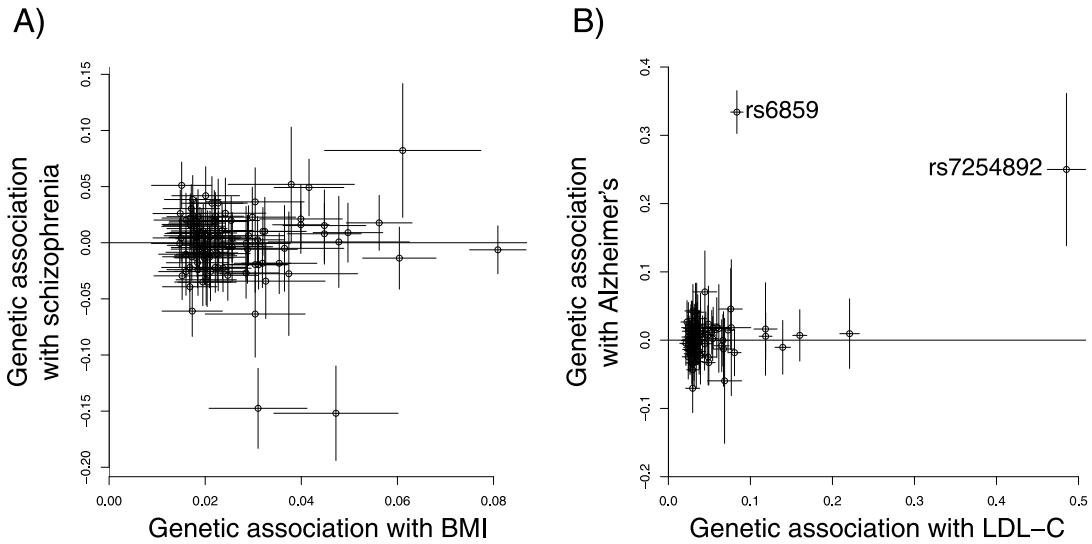


Fig 3. Graph A) displays the estimated genetic associations and 95% confidence intervals with body mass index (BMI, standard deviation units) and with schizophrenia (log odds ratios) for 97 genetic variants. Graph B) displays the estimated genetic associations and 95% confidence intervals with low-density lipoprotein cholesterol (LDL-C, standard deviation units) and with Alzheimer's disease (log odds ratios) for 75 genetic variants: the two outlying variants are labelled with their rsID codes.

https://doi.org/10.1371/journal.pone.0222362.g003

Table 1. Estimates (standard errors) and 95% confidence intervals of the causal effect of body mass index on schizophrenia risk (log odds ratio for schizophrenia per 1 standard deviation increase in body mass index) and low-density lipoprotein cholesterol on Alzheimer's disease risk (log odds ratio for Alzheimer's per 1 standard deviation increase in low-density lipoprotein cholesterol) from the IVW method with: 1) the full set of genetic variants (IVW); 2) robust regression; 3) penalized weights; and 4) robust regression and penalized weights. Results from Lasso penalization with the heterogeneity stopping rule and cross-validation, simple median, weighted median and MR-Egger methods are also presented.

	Estimate (SE)	95% CI
Applied example 1: Causal effect of BMI on schizophrenia risk		
IVW	-0.031 (0.100)	-0.227, 0.165
Robust regression	-0.024 (0.079)	-0.180, 0.132
Penalized weights	-0.056 (0.065)	-0.184, 0.073
Robust regression with penalized weights	-0.052 (0.066)	-0.182, 0.078
Lasso penalization		
Heterogeneity stopping rule	-0.022 (0.055)	-0.131, 0.086
Cross validation	-0.036 (0.087)	-0.207, 0.136
Median		
Simple	-0.073 (0.083)	-0.237, 0.090
Weighted	-0.075 (0.090)	-0.252, 0.102
MR-Egger	0.336 (0.241)	-0.136, 0.808
Applied example 2: Causal effect of LDL-C on AD risk		
IVW	0.239 (0.102)	0.039, 0.439
Robust regression	0.048 (0.038)	-0.027, 0.123
Penalized weights	0.040 (0.042)	-0.043, 0.123
Robust regression with penalized weights	0.046 (0.032)	-0.016, 0.108
Lasso penalization		
Heterogeneity stopping rule	0.032 (0.044)	-0.054, 0.118
Cross validation	0.088 (0.045)	0.000, 0.175
Median		
Simple	0.108 (0.071)	-0.031, 0.247
Weighted	0.046 (0.061)	-0.073, 0.165
MR-Egger	0.391 (0.168)	0.061, 0.722

Abbreviations: SE, standard error; CI, confidence interval; BMI, body mass index; IVW, inverse-variance weighted; LDL-C, low-density lipoprotein cholesterol; AD, Alzheimer's disease.

https://doi.org/10.1371/journal.pone.0222362.t001

schizophrenia; and two outliers in the LDL-C and AD example. The outlying variants (rs6859 and rs7254892) for LDL-C and AD are located near to the *APOE* locus and are associated with AD risk with odds ratios of 1.40 (95% CI: 1.35, 1.44) and 1.28 (95% CI: 1.15, 1.44) respectively [41]. Studentized residuals from the IVW analysis for these variants are 16.5 and −0.95 (all other variants had absolute Studentized residual less than 2), and Cook's distances are 2.51 and 0.11 respectively (all other Cook's distances were less than 0.06).

Estimates and 95% confidence intervals from the Mendelian randomization analyses are provided in Table 1. All of the estimates for BMI and schizophrenia suggest a null causal effect (as also observed in the Mendelian randomization study by Hartwig *et al.* [34]), although there is wide variation in the standard errors. The use of penalized weights and robust regression in the IVW method improved the precision of the estimates. There was little difference in the point estimates or standard errors obtained from the IVW method with penalized weights, and from the IVW method with robust regression and penalized weights. With exception of the IVW and MR-Egger methods, the median estimates were the least precise.

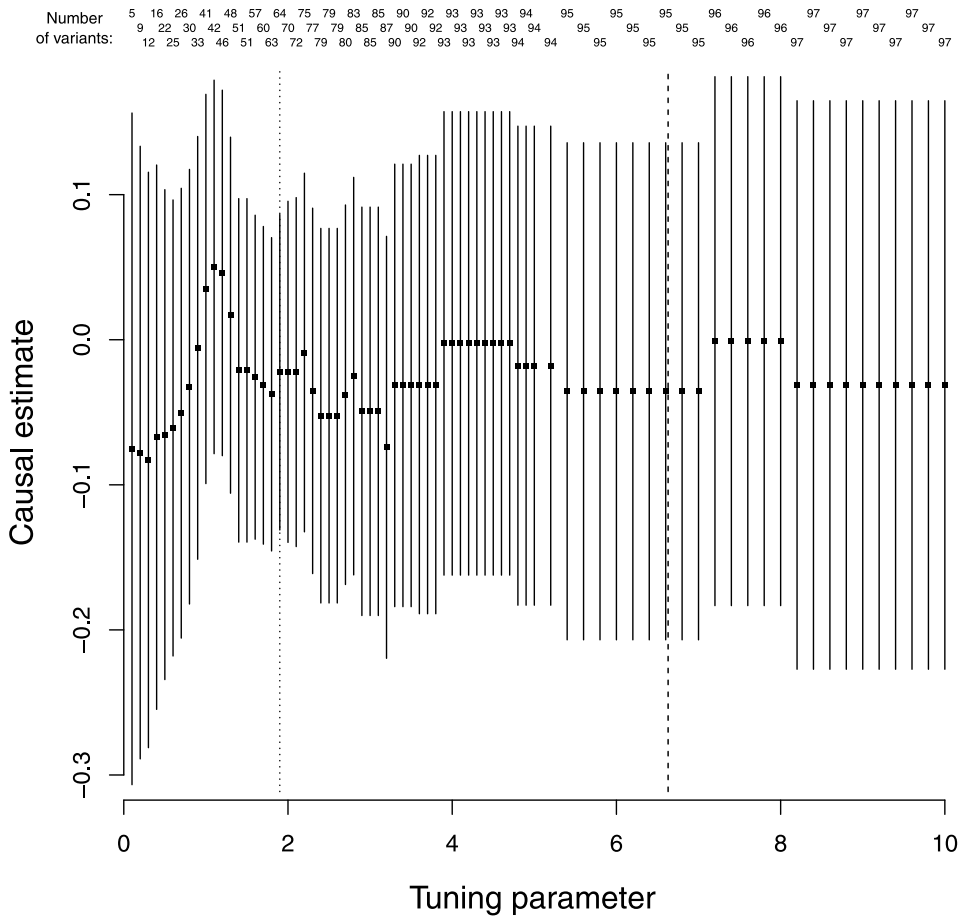


Fig 4. Log odds ratio and 95% confidence intervals for schizophrenia per 1 standard deviation increase in body mass index for different values of the tuning parameter (λ = 0.1, 0.2, . . ., 4.9, 5.0, 5.2, 5.4, . . ., 9.8, 10.0) included in the Lasso regression model. The number of genetic variants included in the IVW models are also displayed. The dotted line at λ = 1.9 is the value of the tuning parameter chosen by the heterogeneity stopping rule. The dashed line at λ = 6.63 is the value chosen by cross-validation.

https://doi.org/10.1371/journal.pone.0222362.g004

The Lasso penalization estimates are displayed in Fig 4 where the causal estimates are relatively similar across the different values of the tuning parameter. The value of the tuning parameter λ was 1.9 under the heterogeneity stopping rule, with 64 genetic variants included in the IVW method. The cross-validation method returned a much larger value of λ = 6.63, with 95 of the 97 variants included in the IVW method.

The estimates from the IVW and MR-Egger methods suggested a positive causal effect of LDL-C on AD risk. This effect was attenuated to the null for the other robust methods. Compared to the robust methods that reported a null causal effect of LDL-C on AD risk, the simple and weighted median estimates had larger standard errors. The estimates from the IVW and MR-Egger methods from Benn *et al.* [40] indicated that lower LDL-C levels may be beneficial

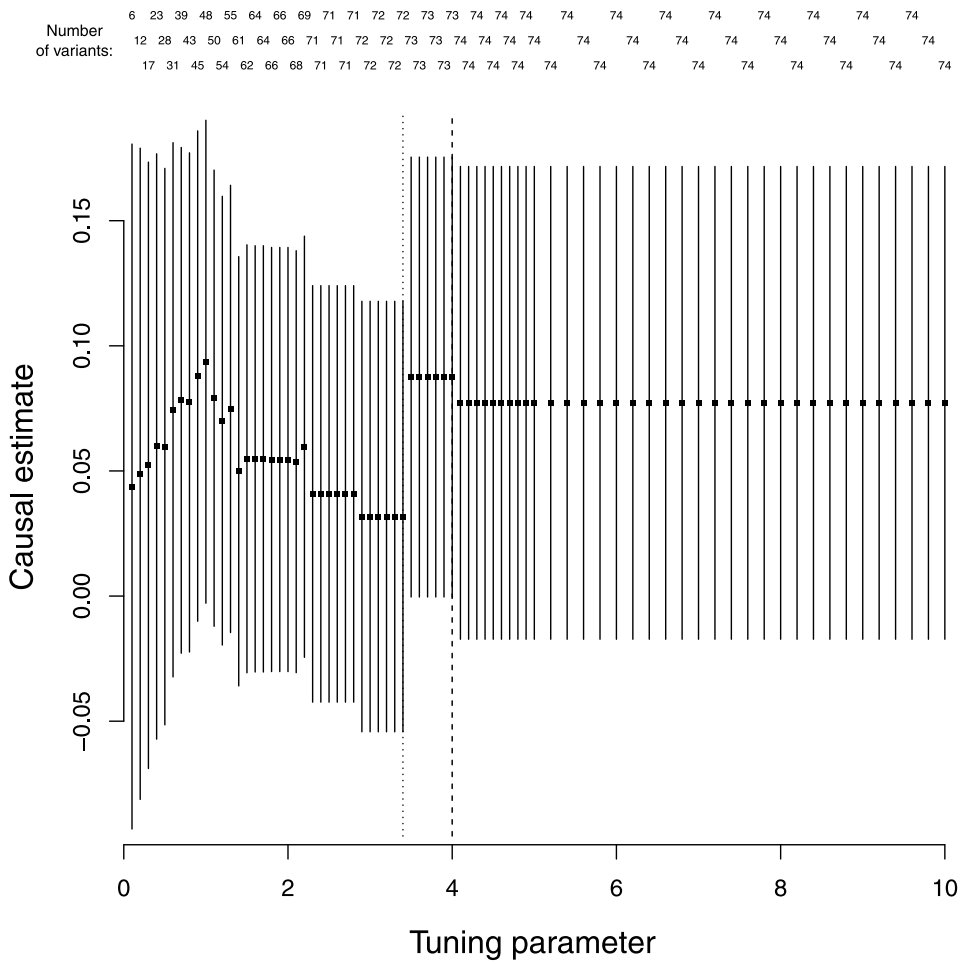


Fig 5. Log odds ratio and 95% confidence intervals for Alzheimer's per 1 standard deviation increase in low-density lipoprotein cholesterol for different values of the tuning parameter (λ = 0.1, 0.2, . . ., 4.9, 5.0, 5.2, 5.4, . . ., 9.8, 10.0) included in the Lasso regression model. The number of genetic variants included in the IVW models are also displayed. The dotted line at λ = 3.4 is the value of the tuning parameter chosen by the heterogeneity stopping rule. The dashed line at λ = 4.00 is the value chosen by cross-validation.

https://doi.org/10.1371/journal.pone.0222362.g005

in reducing AD risk, whereas their estimate from the weighted median method suggested a null effect. Since the genetic variants in the *APOE* gene region tend to be highly pleiotropic [10], it is likely that the positive effects obtained from the IVW models in our analysis and in the paper by Benn *et al.* [40] are driven by these pleiotropic variants, rather than there being a true causal effect of LDL-C on AD risk.

The λ values for the heterogeneity stopping rule (λ = 3.4 based on 72 genetic variants) and cross-validation (λ = 4.00 based on 73 genetic variants) for Lasso penalization were similar (Fig 5). However, the estimate based on 72 genetic variants was much closer to the null,

demonstrating the sensitivity of the IVW method to a single variant. None of the estimates in Fig 5 include information on the rs6859 variant, and this outlying variant was only included in the IVW model for Lasso penalization when $\lambda = 19.8$, whereas the other outlying variant (rs57254892) was included when $\lambda = 3.5$.

The consistency of the results from the robust methods for the BMI and schizophrenia example strengthened the evidence from the primary IVW analysis, providing similar point estimates but with narrower confidence intervals. The LDL-C and AD example highlighted the possibility that only using the IVW method may provide conclusions that are not representative of the majority of the data. Whilst in practice the outlying rs6859 variant could have been identified and removed from the dataset prior to the analysis, the robust approaches identified this outlying variant in an automated manner.

Simulation study

Approaches applied to the simulated data

We applied the approaches introduced in this paper to simulated datasets, including the IVW method with: 1) all the J genetic variants (standard IVW method); 2) robust regression; 3) penalized weights; and 4) robust regression and penalized weights. The Lasso penalization method with the heterogeneity stopping rule was also considered. The bias and coverage properties of the estimates from these robust methods were compared to those from the simple (unweighted) median, weighted median, and MR-Egger methods. Standard errors for the simple and weighted median estimates were obtained through bootstrapping [8]. Robust regression, penalized weights, and robust regression and penalized weights were also applied to the MR-Egger model. The Lasso penalization method was applied to $\lambda = 0.1, 0.2, \ldots, 4.9, 5.0, 5.2, 5.4, \ldots, 9.8, 10.0$ under the heterogeneity stopping rule.

To allow for direct comparisons with the MR-Egger method, and to assess the performance of the methods when the IV assumptions were violated, the simulations followed a similar structure to the simulation study performed in the paper by Bowden *et al.* [8]. The data generating model used in the simulation study is outlined below.

Data generating model

The simulation study generated data in accordance to Fig 6 for participants indexed by $i = 1, \ldots, N$, and genetic variants indexed with $j = 1, \ldots, J$:

$$
\begin{aligned}
U_i &= \sum_{j=1}^{J} \phi_j G_{ij} + \epsilon_{Ui}\,, \\
X_i &= \sum_{j=1}^{J} \beta_{X_j} G_{ij} + U_i + \epsilon_{Xi}\,, \\
Y_i &= \sum_{j=1}^{J} \alpha_j G_{ij} + \theta X_i + U_i + \epsilon_{Yi}\,, \\
G_{ij} &\sim \text{Binomial}(2, 0.3) \text{ independently for all } j = 1, \ldots, J\,, \\
\epsilon_{Ui}, \epsilon_{Xi}, \epsilon_{Yi} &\sim \mathcal{N}(0, 1) \text{ independently}\,,
\end{aligned}
$$

where α_j represents the direct effect of the genetic variant G_j on the outcome, ϕ_j represents the effect of the genetic variant on the confounder U of the risk factor X and outcome Y association, β_{X_j} represents the genetic effect of G_j on X, and θ is the causal effect of X on Y. The error terms ϵ_{Ui}, ϵ_{Xi}, and ϵ_{Yi} were drawn independently from standard normal distributions.

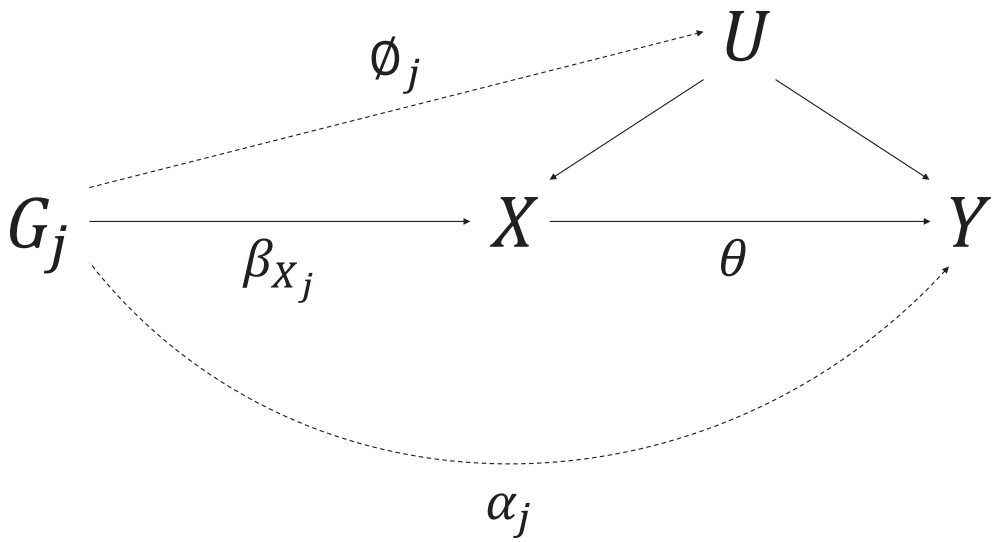


Fig 6. Causal directed acyclic graph used in the data generating model for the simulation study. U represents the set of variables that confound the association between the risk factor X and outcome Y. The genetic effect of G_j on X is β_{X_j}, the direct (pleiotropic) effect of G_j on Y is α_j, the effect of G_j on U is ϕ_j, and the causal effect of X on Y is θ.

https://doi.org/10.1371/journal.pone.0222362.g006

The performance of the robust methods was investigated under a two-sample Mendelian randomization setting with $N = 10,000$ individuals and $J = 15$ genetic variants. Data were generated for $2N$ participants, and the associations of the variants with the risk factor were estimated in the first N participants, and associations with the outcome in the second N participants. Only the summary level data (beta-coefficients and standard errors) were used in the analyses. A one-sample setting was also considered where an additional N participants were simulated and all of the genetic associations were estimated from the same N participants.

If a genetic variant is associated with a confounder of the risk factor–outcome association, then this will affect the variant's association with both the risk factor and the outcome, leading to the violation of the InSIDE assumption. Using this observation, data were simulated to consider the following four scenarios:

- Scenario 1—No pleiotropy, InSIDE automatically satisfied: α_j and ϕ_j were set to zero for all j.
- Scenario 2—Balanced pleiotropy, InSIDE satisfied: $\alpha_j \sim U[0.05, 0.15]$ for invalid variants, with each α_j having a 0.5 probability of being multiplied by -1. ϕ_j was set to zero for all j.
- Scenario 3—Directional pleiotropy, InSIDE satisfied: $\alpha_j \sim U[0.05, 0.15]$ for invalid variants, and ϕ_j was set to zero for all j.
- Scenario 4—Directional pleiotropy, InSIDE violated: $\phi_j \sim U[0.05, 0.10]$ for invalid variants, and α_j was set to zero for all j.

The genetic variants G_j were coded to correspond to a single nucleotide polymorphism with minor allele frequency 0.3. If a genetic variant was a valid IV then α_j and ϕ_j were set to zero in all four scenarios. In Scenarios 2 to 4, the number of invalid IVs was set to 1, 3 and 6.

The causal effect of the risk factor on the outcome was either $\theta = 0$ (null causal effect) or $\theta = 0.3$ (positive causal effect). The effects of the genetic variants on the risk factor (β_{X_j}) were drawn from a uniform distribution between 0.06 and 0.13. 10 000 simulated datasets were generated for each combination of parameters (24 different combinations in total).

Results

The mean proportion of variance in the risk factor explained by the genetic variants (R^2 statistic), mean F statistic, and mean I^2 statistic are contained in Table A in the S2 Appendix for scenarios 1-4 for the null and positive causal effects by the number of invalid instruments. The mean R^2 values were greater than 3% for all of the scenarios, and the minimum mean F-statistic was 20.8. The I^2 statistic ranged from 39.1% to 80.9%. Since violations in the no measurement error (NOME) assumption of the genetic associations with the risk factor can lead to attenuation towards the null for the MR-Egger estimates, and this attenuation is approximately equal to the I^2 statistic, we expected the MR-Egger estimates for the positive causal effect to be severely attenuated towards the null [42].

The number of robust regression models that did not report a standard error (maximum of 2.6% across all of the scenarios considered) are given in Table B in the S2 Appendix. Apart from the calculation of the mean standard error, the robust regression models that did not report a standard error were included in the results, and the power calculations treated the standard error as infinite.

When all of the genetic variants were valid IVs (Table 2), all of the methods produced unbiased estimates of the null causal effect and the Type I error rates were close to the nominal level of 5%. Apart from the simple median method, there was attenuation towards the null with a positive causal effect for all methods, and as expected, this was particularly evident for the MR-Egger method (also observed for Scenarios 2 and 3). Violation of the NOME assumption can lead to inflation of the intercept term in the MR-Egger method [42], and this was true for the simulation study where the power to detect the intercept term for Scenarios 1 and 2 was greater than 5% (Table C in the S2 Appendix). Only 7.5% of the MR-Egger models

Table 2. Mean estimate (mean standard error), standard deviation, coverage of the 95% confidence interval (%), and power at the 5% significance level (%) of the estimates from the IVW model with: 1) the *J* genetic variants (IVW); 2) robust regression; 3) penalized weights; and 4) robust regression and penalized weights for Scenario 1 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect. Results from Lasso penalization with the heterogeneity stopping rule, simple (unweighted) median, weighted median and MR-Egger methods are also provided.

	Null causal effect ($\theta = 0$)				Positive causal effect ($\theta = 0.3$)			
	Estimate (SE)	SD	Cov.	Pow.	Estimate (SE)	SD	Cov.	Pow.
Scenario 1. No pleiotropy, InSIDE automatically satisfied								
IVW	-0.001 (0.061)	0.058	95.7	4.3	0.287 (0.073)	0.069	95.5	98.2
Robust regression	-0.001 (0.066)	0.060	95.1	4.9	0.287 (0.079)	0.072	94.7	94.8
Penalized weights	-0.001 (0.060)	0.059	95.0	5.0	0.289 (0.072)	0.071	94.7	98.2
Robust regression with penalized weights	-0.001 (0.064)	0.061	94.5	5.5	0.288 (0.077)	0.073	94.1	95.7
Lasso penalization	-0.001 (0.060)	0.059	94.8	5.2	0.287 (0.072)	0.071	94.6	98.0
Median								
Simple	-0.002 (0.086)	0.074	97.9	2.1	0.301 (0.105)	0.090	98.0	86.9
Weighted	-0.002 (0.080)	0.071	97.4	2.6	0.277 (0.097)	0.085	96.7	85.7
MR-Egger	-0.001 (0.219)	0.207	96.1	3.9	0.143 (0.261)	0.251	91.0	7.5

Abbreviations: SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse-variance weighted.

https://doi.org/10.1371/journal.pone.0222362.t002

Table 3. Mean estimate (mean standard error), standard deviation, coverage of the 95% confidence interval (%), and power at the 5% significance level (%) of the estimates from the IVW model with: 1) the *J* genetic variants (IVW); 2) robust regression; 3) penalized weights; and 4) robust regression and penalized weights (R and P) for Scenarios 2-4 with a null causal effect ($\theta = 0$) by the number of invalid IVs. Results from Lasso penalization with the heterogeneity stopping rule, simple median, weighted median and MR-Egger methods are also provided.

	1 invalid IV				3 invalid IVs				6 invalid IVs			
	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.
Scenario 2. Balanced pleiotropy, InSIDE satisfied												
IVW	-0.002 (0.089)	0.092	94.7	5.3	0.000 (0.133)	0.136	93.4	6.6	0.000 (0.180)	0.183	93.0	7.0
Robust	-0.002 (0.069)	0.065	94.3	5.7	0.000 (0.096)	0.087	94.5	5.5	0.001 (0.196)	0.173	94.3	5.6
Penalized	-0.002 (0.062)	0.064	94.2	5.8	0.000 (0.066)	0.077	91.1	8.9	0.001 (0.075)	0.116	81.5	18.5
R and P	-0.002 (0.071)	0.065	94.6	5.4	0.001 (0.094)	0.078	94.7	5.2	0.001 (0.160)	0.119	91.5	7.3
Lasso	-0.002 (0.063)	0.065	94.4	5.6	0.000 (0.071)	0.080	91.7	8.3	0.001 (0.088)	0.129	84.5	15.5
Median												
Simple	-0.002 (0.090)	0.080	97.4	2.6	0.001 (0.097)	0.094	96.5	3.5	0.002 (0.115)	0.132	92.7	7.3
Weighted	-0.001 (0.082)	0.076	96.9	3.2	0.000 (0.089)	0.090	95.2	4.8	0.000 (0.101)	0.133	88.9	11.1
MR-Egger	-0.004 (0.317)	0.335	92.7	7.3	-0.009 (0.477)	0.496	92.7	7.3	-0.006 (0.646)	0.661	93.0	7.0
Scenario 3. Directional pleiotropy, InSIDE satisfied												
IVW	0.064 (0.089)	0.064	94.8	5.2	0.194 (0.126)	0.076	76.0	24.0	0.388 (0.154)	0.089	16.1	83.9
Robust	0.010 (0.069)	0.064	94.3	5.7	0.069 (0.113)	0.083	93.9	6.1	0.335 (0.227)	0.105	63.6	36.4
Penalized	0.007 (0.062)	0.063	94.2	5.8	0.033 (0.067)	0.078	89.2	10.8	0.148 (0.082)	0.137	57.3	42.7
R and P	0.005 (0.072)	0.065	94.8	5.2	0.025 (0.092)	0.079	93.2	6.7	0.115 (0.138)	0.147	78.6	20.9
Lasso	0.006 (0.063)	0.065	94.2	5.8	0.031 (0.071)	0.080	90.3	9.7	0.164 (0.096)	0.146	60.5	39.5
Median												
Simple	0.021 (0.089)	0.077	97.3	2.7	0.074 (0.100)	0.086	92.9	7.2	0.224 (0.134)	0.124	64.3	35.7
Weighted	0.017 (0.082)	0.074	96.9	3.1	0.065 (0.090)	0.085	91.7	8.3	0.210 (0.110)	0.149	56.9	43.1
MR-Egger	-0.003 (0.318)	0.334	92.9	7.2	-0.001 (0.450)	0.465	93.2	6.8	-0.004 (0.544)	0.562	92.4	7.6
Scenario 4. Directional pleiotropy, InSIDE violated												
IVW	0.077 (0.070)	0.058	83.4	16.7	0.186 (0.075)	0.056	25.6	74.4	0.290 (0.071)	0.050	0.3	99.7
Robust	0.031 (0.085)	0.069	93.9	6.0	0.142 (0.127)	0.082	73.1	26.0	0.289 (0.079)	0.053	3.3	96.6
Penalized	0.021 (0.061)	0.070	89.2	10.8	0.083 (0.063)	0.091	64.1	35.9	0.231 (0.061)	0.092	12.2	87.8
R and P	0.018 (0.071)	0.070	92.7	7.3	0.075 (0.084)	0.095	76.2	23.7	0.230 (0.074)	0.101	19.0	80.8
Lasso	0.024 (0.062)	0.073	88.2	11.8	0.116 (0.066)	0.099	51.1	48.9	0.286 (0.066)	0.070	2.2	97.8
Median												
Simple	0.020 (0.089)	0.077	97.3	2.7	0.071 (0.092)	0.083	89.9	10.1	0.192 (0.088)	0.091	40.0	60.0
Weighted	0.055 (0.082)	0.077	91.0	9.0	0.198 (0.081)	0.097	34.8	65.2	0.343 (0.069)	0.074	0.5	99.5
MR-Egger	0.305 (0.214)	0.219	66.8	33.2	0.539 (0.197)	0.183	21.3	78.7	0.644 (0.182)	0.165	5.1	94.9

Abbreviations: IV, instrumental variable; Est. estimate; SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted method; R and P, robust and penalized.

https://doi.org/10.1371/journal.pone.0222362.t003

detected a positive causal effect, and apart from the median estimators, all of the robust methods had approximately 95% power to detect the positive causal effect.

Although the mean estimates in Scenario 2 (Tables 3 and 4) were similar to those in Scenario 1, there were clear differences in the precision of the estimates for the null and positive causal effects, with most of the methods reporting larger mean standard errors under Scenario 2. The mean standard error increased as the number of invalid instruments increased for all methods. The IVW model with penalized weights had the most precise estimates, but suffered from inflated Type I error rates and poor coverage. The simple and weighted median estimators performed just as well, if not better, than the other robust methods for Scenario 2.

Table 4. Mean estimate (mean standard error), standard deviation, coverage of the 95% confidence interval (%), and power at the 5% significance level (%) of the estimates from the IVW model with: 1) the *J* genetic variants (IVW); 2) robust regression; 3) penalized weights; and 4) robust regression and penalized weights (R and P) for Scenarios 2-4 with a positive causal effect ($\theta = 0.3$) by the number of invalid IVs. Results from the Lasso penalization method with the heterogeneity stopping rule, simple median, weighted median and MR-Egger methods are also provided.

	1 invalid IV				3 invalid IVs				6 invalid IVs			
	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.	Est. (SE)	SD	Cov.	Pow.
Scenario 2. Balanced pleiotropy, InSIDE satisfied												
IVW	0.286 (0.097)	0.100	94.0	80.9	0.288 (0.139)	0.140	93.5	54.5	0.285 (0.184)	0.184	93.3	34.3
Robust	0.287 (0.084)	0.079	93.9	91.0	0.288 (0.116)	0.107	93.6	71.1	0.286 (0.193)	0.178	93.8	34.5
Penalized	0.289 (0.074)	0.079	93.2	96.5	0.291 (0.080)	0.098	88.6	91.8	0.295 (0.090)	0.147	78.4	80.5
R and P	0.289 (0.083)	0.080	93.5	92.5	0.290 (0.100)	0.097	92.2	81.6	0.295 (0.145)	0.147	88.1	59.4
Lasso	0.287 (0.076)	0.080	93.2	95.4	0.288 (0.085)	0.102	89.4	88.0	0.288 (0.108)	0.167	80.5	69.9
Median												
Simple	0.302 (0.109)	0.097	97.5	83.1	0.302 (0.118)	0.113	96.1	75.3	0.303 (0.136)	0.155	92.5	61.4
Weighted	0.276 (0.100)	0.091	96.1	81.9	0.277 (0.106)	0.108	94.0	75.6	0.277 (0.119)	0.152	88.2	63.2
MR-Egger	0.143 (0.349)	0.363	90.2	9.0	0.138 (0.495)	0.518	91.2	8.4	0.126 (0.657)	0.681	92.1	7.5
Scenario 3. Directional pleiotropy, InSIDE satisfied												
IVW	0.353 (0.098)	0.075	96.1	97.8	0.482 (0.133)	0.087	81.5	99.3	0.673 (0.160)	0.101	28.3	100
Robust	0.306 (0.084)	0.077	95.1	94.8	0.383 (0.134)	0.099	93.9	86.2	0.631 (0.205)	0.112	60.8	90.8
Penalized	0.303 (0.074)	0.078	93.8	98.0	0.346 (0.081)	0.100	86.4	97.8	0.511 (0.098)	0.164	47.6	99.1
R and P	0.300 (0.083)	0.080	94.1	93.5	0.335 (0.102)	0.102	91.3	88.8	0.485 (0.142)	0.179	66.2	86.9
Lasso	0.301 (0.076)	0.079	94.0	97.4	0.340 (0.086)	0.104	88.2	96.5	0.513 (0.113)	0.168	53.8	98.5
Median												
Simple	0.329 (0.110)	0.095	97.5	89.5	0.393 (0.125)	0.108	93.5	93.2	0.572 (0.158)	0.150	61.8	97.2
Weighted	0.300 (0.100)	0.090	97.2	88.1	0.356 (0.111)	0.104	94.5	93.0	0.516 (0.131)	0.161	63.6	97.4
MR-Egger	0.142 (0.345)	0.353	90.9	8.7	0.138 (0.468)	0.485	91.3	8.1	0.137 (0.555)	0.576	91.9	8.0
Scenario 4. Directional pleiotropy, InSIDE violated												
IVW	0.367 (0.080)	0.071	88.8	99.8	0.478 (0.084)	0.067	42.2	100	0.582 (0.078)	0.062	2.3	100
Robust	0.329 (0.100)	0.082	94.3	90.2	0.447 (0.128)	0.085	73.7	90.2	0.581 (0.087)	0.066	8.2	99.7
Penalized	0.323 (0.072)	0.086	88.6	98.2	0.403 (0.072)	0.102	61.2	98.9	0.546 (0.068)	0.092	11.2	99.9
R and P	0.318 (0.085)	0.087	92.4	94.1	0.397 (0.095)	0.107	72.7	93.6	0.547 (0.077)	0.098	16.0	98.4
Lasso	0.323 (0.073)	0.089	87.9	97.6	0.430 (0.076)	0.105	52.7	99.3	0.579 (0.073)	0.079	4.4	100
Median												
Simple	0.328 (0.108)	0.095	97.0	89.9	0.387 (0.111)	0.101	90.0	95.1	0.509 (0.101)	0.101	43.4	99.5
Weighted	0.344 (0.099)	0.095	94.0	94.4	0.496 (0.097)	0.108	47.2	99.7	0.625 (0.085)	0.087	3.4	100
MR-Egger	0.488 (0.254)	0.259	86.1	51.8	0.767 (0.233)	0.220	45.4	90.0	0.887 (0.214)	0.197	20.2	98.1

Abbreviations: IV, instrumental variable; Est. estimate; SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted method; R and P, robust and penalized.

https://doi.org/10.1371/journal.pone.0222362.t004

In Scenario 3 (directional pleiotropy, InSIDE satisfied), the IVW method produced biased causal estimates with inflated Type I error rates, and the degree of bias increased with the number of invalid IVs. With one invalid instrument, estimates from the robust methods were only slightly biased and Type I error rates were fairly well controlled. As the number of instruments increased, bias in the estimates for the robust methods also increased, although the magnitude of bias was smaller than the IVW method, and Type I error inflation was less severe. Robust regression with penalized weights performed reasonably well when there was 1 or 3 invalid instruments. Although the median methods give unbiased estimates asymptotically (that is, as the number of participants increases), when pleiotropic effects are directional there is some bias with a finite sample.

In Scenario 4 (directional pleiotropy, InSIDE violated), all of the robust methods produced biased estimates. When there were only one invalid instrument, the magnitude of bias from the robust methods was less severe than the IVW method, and this was particularly true for robust regression with penalized weights. As the number of invalid IVs increased, the performance of the robust methods worsened, and there was little advantage in applying the robust methods compared to the median estimator in Scenario 4 when 6 of the 15 genetic variants were invalid IVs. In this scenario, bias is greater for the weighted median than for the simple median method as the invalid genetic variants are on average more strongly associated with the risk factor than the valid ones. This is because invalid variants are associated with the risk factor directly and via their effect on the confounder. In practical applications, invalid genetic variants will not necessarily be more strongly associated with the risk factor than valid ones, and so the simple median will not necessarily perform better than the weighted median method.

While results were fairly similar for most of the methods, results from the MR-Egger method were often quite different. This is because the other methods are fairly similar in their assumptions (that most genetic variants are valid IVs) and their mode of operation (variants with causal estimates that differ from the consensus are penalized or down-weighted). This highlights the importance in an applied analysis of performing a range of methods that make different assumptions, rather than multiple methods that make similar assumptions [43].

Results from applying robust regression and penalized weights to the MR-Egger method are provided in Table D in the S2 Appendix. Although we had hoped that the combination of the MR-Egger method and approaches to reduce the influence of outlying variants would be synergistic in improving robustness, findings were disappointing, and all of the models were affected by the violation of the NOME assumption. A reason for this is the flexibility of the method: in allowing the intercept to differ from zero and allowing outliers that deviate from the regression model, the method permits the IV assumptions to be violated in quite a broad way. In a substantial number of cases, the method identified the wrong variants as invalid, finding an incorrect configuration of valid and invalid variants that appeared to fit the data better.

Finally, results from the one-sample setting are provided in Tables E and F in the S2 Appendix. Bias in the direction of the observational association was observed for all methods. As with the two–sample setting, the median estimators and robust regression with penalized weights produced the least biased estimates, and the IVW with penalized weights was the most precise.

Increased number of genetic variants

Since many of the methods described in this paper are based on asymptotic theory, it was anticipated that there would be an improvement in the performance of the methods when the data were generated with a larger number of genetic variants. We therefore repeated the simulation study for Scenarios 2–4 for 1 000 simulated datasets with the number of genetic variants increased from 15 to 100, and the number of invalid IVs increased from 1, 3 and 6 to 5, 15 and 30. The bounds of the uniform distribution used to generate the genetic associations with the risk factor (β_{X_j}) were multiplied by $\sqrt{\frac{15}{100}}$ to ensure the average R^2 values were comparable with the original simulation study. The IVW model with: 1) the full set of genetic variants; 2) robust regression; 3) penalized weights; and 4) robust regression and penalized weights were all applied to the dataset. The Lasso penalization method with the heterogeneity stopping rule was also considered.

Results. The mean R^2 statistic, F-statistic, and I^2 statistic are contained in Table G in the S2 Appendix for Scenarios 2–4 for the null and positive causal effect by the number of invalid IVs. The mean R^2 values for the 100 genetic variants were slightly higher than the values reported in the original simulation study (Table A in the S2 Appendix). For all of the scenarios considered, there was a significant reduction in the mean F-statistic and I^2 statistic, and we therefore expected the estimates to be affected by weak instrument bias.

Results from the simulation study for the IVW model with: 1) the J genetic variants (IVW); 2) robust regression; 3) penalized weights; and 4) robust regression and penalized weights, and the Lasso penalization method with the heterogeneity stopping rule are provided in Table 5.

The reduction in the strength of the IVs led to weak instrument bias, and there was severe attenuation towards the null for the positive causal effect (Table 5). For the null causal effect, there was little difference in the performance of the robust methods with the increased number of genetic variants. In fact, the methods performed worst under Scenario 4 when 100 variants were included in the data generating model rather than 15 (Table 5). Due to the attenuation of the positive causal effect when the number of variants was increased to 100, it was difficult to compare the results to the original simulations. Nevertheless, there was no evidence to suggest that the performances of the robust methods improved when the number of genetic variants was increased.

Discussion

In this paper, we have introduced three robust approaches for Mendelian randomization with summary level data that downweight the influence of heterogeneous causal estimates. The applied examples considered in this paper illustrate the importance of using a variety of methods in a Mendelian randomization analysis. The results from the robust methods support a null causal effect of BMI on schizophrenia risk. While the IVW and MR-Egger methods produced positive estimates that were strongly influenced by pleiotropic variants in the *APOE* gene region, the proposed methods were able to give null estimates that were unaffected by these outlying variants.

We also performed a simulation study to compare the robust approaches to the IVW, simple median, weighted median, and MR-Egger methods. The simulation study highlighted the sensitivity of the IVW method to violations in the IV assumptions, and the requirement for robust methods to be considered in the sensitivity analysis of a Mendelian randomization study. The simulations also demonstrated the impact of violating the NOME assumption on the estimates from the MR-Egger methods. Since it was not feasible to adjust for the violation of the NOME assumption through the SIMEX method [42] in the simulation study for computational reasons, it was difficult to compare the performance of the robust methods to MR-Egger.

Robust regression with penalized weights consistently produced the least biased estimates in the simulation study. Although the power and bias of this approach was significantly better than the standard IVW method when the IV assumptions were violated, it suffered from poor coverage and increased Type I error rates, particularly when there was a high proportion of invalid instruments. When there was only one invalid instrument, robust regression with penalized weights produced more precise estimates than the median estimator. However, as the number of invalid instruments increased there was little advantage of using robust regression with penalized weights compared to the median estimator.

Interpretation of heterogeneity among the causal ratio estimates

Throughout this paper, we have assumed that heterogeneity of the causal ratio estimates is indicative of violations in the IV assumptions, particularly the presence of pleiotropic effects.

Table 5. Results from the simulation study when 100 genetic variants were simulated for 1 000 datasets. Mean estimate (mean standard error), standard deviation, coverage of the 95% confidence interval (%), and power at the 5% significance level (%) of the estimates from the IVW model with: 1) the *J* genetic variants (IVW); 2) robust regression; 3) penalized weights; and 4) robust regression and penalized weights (R and P) for Scenarios 2-4 with a null causal effect ($\theta = 0$) and positive causal effect ($\theta = 0.3$) by the number of invalid instrumental variables. Results from the Lasso penalization method with the heterogeneity stopping rule are also presented.

	5 invalid IV				15 invalid IVs				30 invalid IVs			
	Est. (SE)	**SD**	**Cov.**	**Pow.**	**Est. (SE)**	**SD**	**Cov.**	**Pow.**	**Est. (SE)**	**SD**	**Cov.**	**Pow.**
Null causal effect ($\theta = 0$)												
Scenario 2. Balanced pleiotropy, InSIDE satisfied												
IVW	-0.003 (0.072)	0.071	95.0	5.0	-0.003 (0.103)	0.105	94.9	5.1	0.000 (0.138)	0.144	94.0	6.0
Robust	-0.001 (0.054)	0.051	95.8	4.2	-0.001 (0.065)	0.066	93.7	6.3	0.005 (0.115)	0.114	95.7	4.3
Penalized	-0.001 (0.051)	0.051	94.8	5.2	-0.001 (0.054)	0.063	91.3	8.7	0.001 (0.060)	0.081	86.3	13.7
R and P	-0.001 (0.055)	0.052	95.8	4.2	-0.001 (0.064)	0.062	95.6	4.4	0.000 (0.087)	0.078	96.5	3.5
Lasso	-0.001 (0.051)	0.052	94.2	5.8	-0.002 (0.055)	0.063	91.5	8.5	0.002 (0.062)	0.081	87.6	12.4
Scenario 3. Directional pleiotropy, InSIDE satisfied												
IVW	0.096 (0.071)	0.058	77.3	22.7	0.287 (0.099)	0.070	9.0	91.0	0.572 (0.126)	0.088	0.0	100
Robust	0.014 (0.055)	0.053	94.8	5.2	0.070 (0.072)	0.064	87.0	13.0	0.355 (0.169)	0.103	41.3	58.7
Penalized	0.012 (0.051)	0.053	93.9	6.1	0.043 (0.054)	0.061	83.9	16.1	0.156 (0.062)	0.094	34.3	65.7
R and P	0.009 (0.055)	0.054	95.2	4.7	0.031 (0.064)	0.061	91.7	8.3	0.108 (0.087)	0.093	74.6	25.4
Lasso	0.013 (0.051)	0.054	93.2	6.8	0.044 (0.055)	0.062	83.4	16.6	0.165 (0.063)	0.095	32.6	67.4
Scenario 4. Directional pleiotropy, InSIDE violated												
IVW	0.170 (0.052)	0.046	7.2	92.8	0.349 (0.049)	0.043	0.0	100	0.476 (0.043)	0.036	0.0	100
Robust	0.076 (0.079)	0.065	87.2	12.8	0.310 (0.089)	0.057	7.7	92.1	0.475 (0.048)	0.038	0.0	100
Penalized	0.053 (0.049)	0.064	72.9	27.1	0.187 (0.046)	0.082	13.5	86.5	0.401 (0.040)	0.062	0.0	100
R and P	0.047 (0.058)	0.064	84.5	15.5	0.184 (0.065)	0.087	26.7	73.3	0.409 (0.044)	0.062	0.0	100
Lasso	0.072 (0.048)	0.068	62.9	37.1	0.276 (0.043)	0.072	0.7	99.3	0.474 (0.035)	0.051	0.0	100
Positive causal effect ($\theta = 0.3$)												
Scenario 2. Balanced pleiotropy, InSIDE satisfied												
IVW	0.227 (0.079)	0.076	17.8	82.2	0.229 (0.108)	0.113	45.5	54.5	0.228 (0.141)	0.136	64.2	35.8
Robust	0.227 (0.065)	0.062	6.1	93.9	0.233 (0.080)	0.082	18.6	81.4	0.230 (0.126)	0.119	54.3	45.7
Penalized	0.230 (0.061)	0.062	3.9	96.1	0.241 (0.064)	0.079	8.2	91.8	0.241 (0.071)	0.100	15.7	84.3
R and P	0.229 (0.064)	0.063	4.8	95.2	0.237 (0.072)	0.077	11.1	88.9	0.236 (0.088)	0.097	25.5	74.5
Lasso	0.227 (0.061)	0.064	4.8	95.2	0.232 (0.065)	0.079	9.2	90.8	0.230 (0.072)	0.095	17.8	82.2
Scenario 3. Directional pleiotropy, InSIDE satisfied												
IVW	0.323 (0.079)	0.068	1.1	98.9	0.514 (0.107)	0.081	0	100	0.804 (0.133)	0.098	0.0	100
Robust	0.251 (0.066)	0.066	2.9	97.1	0.342 (0.093)	0.081	1.3	98.7	0.654 (0.162)	0.107	0.2	99.8
Penalized	0.251 (0.061)	0.066	2.1	97.9	0.308 (0.065)	0.080	0.8	99.2	0.490 (0.076)	0.121	0.0	100
R and P	0.246 (0.065)	0.067	4.1	95.9	0.291 (0.074)	0.080	3.2	96.8	0.442 (0.102)	0.123	1.6	98.3
Lasso	0.247 (0.061)	0.066	2.5	97.5	0.297 (0.065)	0.081	1.3	98.7	0.463 (0.074)	0.115	0.0	100
Scenario 4. Directional pleiotropy, InSIDE violated												
IVW	0.411 (0.061)	0.060	0.0	100	0.609 (0.058)	0.054	0.0	100	0.747 (0.051)	0.045	0.0	100
Robust	0.327 (0.095)	0.076	4.5	95.5	0.575 (0.093)	0.065	0.6	99.4	0.746 (0.057)	0.047	0.0	100
Penalized	0.307 (0.058)	0.080	0.9	99.1	0.479 (0.053)	0.093	0.1	99.9	0.689 (0.046)	0.066	0.0	100
R and P	0.298 (0.072)	0.080	1.7	98.3	0.478 (0.073)	0.098	0.2	99.6	0.697 (0.051)	0.066	0.0	100
Lasso	0.314 (0.057)	0.08	0.3	99.7	0.544 (0.050)	0.081	0.0	100	0.742 (0.041)	0.061	0.0	100

Abbreviations: IV, instrumental variable; Est. estimate; SE, standard error; SD, standard deviation; Cov., coverage; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted; R and P, robust regression and penalized weights.

https://doi.org/10.1371/journal.pone.0222362.t005

However, heterogeneity among the causal ratio estimates may arise for a number of reasons [44]. For example, there may be multiple mechanisms of intervention on a complex risk factor, each of which has an associated causal effect. For a two-sample Mendelian randomization analysis, there may be heterogeneity among the causal ratio estimates due to substantial differences in the study populations used to estimate the genetic associations with the risk factor and outcome. The robust approaches considered in this paper penalize genetic variants with heterogeneous causal ratio estimates regardless of how this heterogeneity has materialised. As such, these methods should only be employed if it is suspected that the IV assumptions have been violated, and other possible reasons for heterogeneity among the causal ratio estimates explored.

Issues with penalizing genetic variants

The simulation study has highlighted some of the disadvantages of excluding or downweighting genetic variants from Mendelian randomization analyses. Excluding genetic variants with heterogeneous causal estimates will generally reduce the standard error of the estimate. However, too much penalization can potentially result in artificial overconfidence in the precision of the causal estimate, leading to poor coverage of the true causal effect and increased Type I error rates, as seen for Lasso penalization. If the excluded genetic variants are truly invalid IVs then removing them from the analysis will reduce bias and improve the precision of the causal estimate. However, outlying or heterogeneous causal ratio estimates may be valid IVs, and so removing them from the analysis would be inappropriate. On balance, it may be more appropriate to consider approaches that reduce the contribution that heterogeneous ratio estimates have on the causal estimate, such as the median estimator or robust regression, rather than excluding them from the analysis. If a large number of variants are identified as outliers, then researchers should consider reporting that the Mendelian randomization analysis is inconclusive, rather than reporting a causal estimate.

Implication for Mendelian randomization studies

The purpose of this paper was not to promote one robust method for Mendelian randomization over another, but to emphasize the need for multiple sensitivity analyses that make different sets of assumptions. Although we acknowledge that none of the proposed methods performed significantly better than the median estimator, the extensions proposed in this paper should provide additional confidence in the findings from a conventional Mendelian randomization analysis, particularly when the causal estimates are consistent. Genetic variants that are downweighted or excluded from the analysis by the robust methods should be examined for pleiotropy to determine whether they should be removed from the dataset. The methods proposed here are likely to be useful for Mendelian randomization analyses performed for large numbers of risk factors in an automated manner, such as for -omics risk factors measured on a high-throughput platform. These methods can help a researcher rapidly triage whether a positive causal estimate from the standard IVW method is evidenced just by a small number of variants (as in the LDL-cholesterol and Alzheimer's disease example), or by the majority of variants.

The methods introduced in this paper, particularly robust regression with penalized weights, may be more suited to certain scenarios than the median estimator. In the applied example for LDL-C and AD risk, there were two variants that appeared to be clear outliers. The median estimator and robust regression with penalized weights both suggested that there was a null causal effect of LDL-C on AD risk, but the estimates from the median estimator were less precise. This observation of robust regression producing more precise estimates was

also observed in the simulation study when there was one invalid IV. Robust regression with penalized weights may be a useful addition to sensitivity analyses in Mendelian randomization when there are a small proportion of variants with heterogeneous causal estimates.

Limitations

We found that the Lasso penalization method may be more appropriate in an applied setting, where the estimates can be reported over a range of values of the tuning parameter. The practicality of applying Lasso penalization to the simulation study was more restrictive, and required an automated approach to selecting the tuning parameter.

Whilst we appreciate the limitation of only considering methods with uncorrelated genetic variants, we argue that robust methods should be used when the IV assumptions are in doubt, and therefore using one genetic variant from each gene region is a sensible (although conservative) approach for robust methods in an applied Mendelian randomization analysis. This is because including multiple variants from a single region may mean that region receives a disproportionate weight in the analysis, and so the validity of the analysis would be overly dependent on the validity of these variants. This could be problematic as correlated variants are likely to all be valid or all be invalid, particularly if they are all in the same gene region. If an analyst does want to include correlated variants in an analysis, this can be done by first calculating the appropriate weighting matrix based on the inverse-variance weights and the correlations between variants, and multiplying the genetic associations by the Cholesky decomposition of this matrix, as described previously [45]. Software code to do this is provided in S1 Appendix. However, we caution that no allowance is made that correlated variants are likely to all be valid or all invalid simultaneously, as the methods treat all association estimates as separate datapoints.

The violation of the NOME assumption limited the utility of the simulation study as the estimates from MR-Egger could not be compared to the robust methods. Given that MR-Egger is frequently used as part of a sensitivity analysis in Mendelian randomization studies, this could be viewed as a weakness of the simulation study.

The main simulation study was also limited by the number of genetic variants considered in the data generating model. Since GWASs are now being performed on large study populations, and estimates of genetic associations are publicly available from large consortia, only considering 15 variants in the simulation study may have been conservative. We tried to rectify this limitation by re-performing the simulation study with 100 genetic variants (but keeping the overall R^2 statistic similar) and found that there was significant attenuation towards the null due to weak instrument bias. We had thought that the performances of some of the robust methods would have improved by increasing the number of genetic variants as the methods are based on asymptotic theory. However, we did not find any significant improvements in the methods, and in some cases, the performance of the models worsened with the increased number of genetic variants.

Conclusion

This paper has highlighted the difficulty in robust causal inference when genetic variants in a Mendelian randomization analysis violate the IV assumptions. The extensions proposed in this paper are by no means perfect; even when a small proportion of the variants were invalid IVs, all methods had inflated Type I error rates in at least one scenario. Nevertheless, the Type I error rate for the proposed extensions was substantially better than the IVW method and MR-Egger when the InSIDE assumption was violated.

This paper has demonstrated the benefits of using multiple robust methods as part of a sensitivity analysis. We suggest that the IVW method using robust regression with penalized weights may be a worthwhile additional sensitivity analysis to be performed in a Mendelian randomization analysis in addition to previously proposed methods.

Supporting information

S1 Appendix. Software code. R code for performing the approaches outlined in the paper, and extracting genetic association estimates.
(PDF)

S2 Appendix. Supplementary tables from the simulation study. Additional results from the simulation study.
(PDF)

Acknowledgments

We would like to thank Dr Jack Bowden (University of Bristol) for comments made on an earlier version of this manuscript.

Author Contributions

Conceptualization: Stephen Burgess.

Data curation: Stephen Burgess.

Formal analysis: Jessica M. B. Rees.

Investigation: Angela M. Wood, Frank Dudbridge.

Methodology: Jessica M. B. Rees, Stephen Burgess.

Supervision: Stephen Burgess.

Writing – original draft: Jessica M. B. Rees, Stephen Burgess.

Writing – review & editing: Angela M. Wood, Frank Dudbridge.

References

1. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology. 2003; 32(1):1–22.
2. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine. 2008; 27 (8):1133–1163. https://doi.org/10.1002/sim.3034 PMID: 17886233
3. Greenland S. An introduction to instrumental variables for epidemiologists. International Journal of Epidemiology. 2000; 29(4):722–729. https://doi.org/10.1093/ije/29.4.722 PMID: 10922351
4. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. European Journal of Epidemiology. 2015; 30(7):543–552. https://doi.org/10.1007/s10654-015-0011-z PMID: 25773750
5. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genetic Epidemiology. 2013; 37(7):658–665. https://doi.org/10.1002/gepi.21758 PMID: 24114802
6. Angrist JD, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. Journal of the American Statistical Association. 1995; 90(430):431–442. https://doi.org/10.1080/01621459.1995.10476535

7. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. International Journal of Epidemiology. 2013; 42(4):1134–1144. https://doi.org/10.1093/ije/dyt093 PMID: 24062299
8. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic Epidemiology. 2016; 40(4):304–314. https://doi.org/10.1002/gepi.21965 PMID: 27061298
9. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology. 2015; 44 (2):512–525. https://doi.org/10.1093/ije/dyv080 PMID: 26050253
10. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nature Genetics. 2018; 50(5):693–698. https://doi.org/10.1038/s41588-018-0099-7 PMID: 29686387
11. Dai JY, Peters U, Wang X, Kocarnik J, Chang-Claude J, Slattery ML, et al. Diagnostics for Pleiotropy in Mendelian Randomization Studies: Global and Individual Tests for Direct Effects. American Journal of Epidemiology. 2018; 187(12):2672–2680. https://doi.org/10.1093/aje/kwy177 PMID: 30188971
12. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. Nature Communications. 2018; 9 (1):224. https://doi.org/10.1038/s41467-017-02317-2 PMID: 29335400
13. Del Greco M F, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Statistics in Medicine. 2015; 34(21):2926–2940. https://doi.org/10.1002/sim.6522
14. Bowden J, Del Greco M F, Minelli C, Zhao Q, Lawlor DA, Sheehan NA, et al. Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. International Journal of Epidemiology. 2018.
15. Huber P. Robust Statistics. Wiley; 2009.
16. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. International Journal of Epidemiology. 2017. https://doi.org/10.1093/ije/dyx034 PMID: 28398548
17. Burgess S, CCGC (CHD CRP Genetics Collaboration). Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. Statistics in Medicine. 2013; 32 (27):4726–4747. https://doi.org/10.1002/sim.5871 PMID: 23733419
18. Vansteelandt S, Bowden J, Babanezhad M, Goetghebeur E. On instrumental variables estimation of causal odds ratios. Statistical Science. 2011; 26(3):403–422. https://doi.org/10.1214/11-STS360
19. Kolesár M, Chetty R, Friedman J, Glaeser E, Imbens G. Identification and inference with many invalid instruments. Journal of Business & Economic Statistics. 2015; 33(4):474–484. https://doi.org/10.1080/07350015.2014.978175
20. Koller M, Stahel W. Sharpening wald-type inference in robust regression for small samples. Computational Statistics & Data Analysis. 2011; 55(8):2504–2515. https://doi.org/10.1016/j.csda.2011.02.014
21. Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, et al. robustbase: Basic Robust Statistics; 2015 URL http://cran.r-project.org/package=robustbase, R package version 0.92-5.
22. Bowden J, Hemani G, Davey Smith G. Invited Commentary: Detecting Individual and Global Horizontal Pleiotropy in Mendelian Randomization—A Job for the Humble Heterogeneity Statistic? American Journal of Epidemiology. 2018; 187(12):2681–2685. https://doi.org/10.1093/aje/kwy185 PMID: 30188969
23. Rucker G, Schwarzer G, Carpenter JR, Binder H, Schumacher M. Treatment effect estimates adjusted for small-study effects via a limit meta-analysis. Biostatistics. 2011; 12(1):122–142. https://doi.org/10.1093/biostatistics/kxq046 PMID: 20656692
24. Kang H, Zhang A, Cai T, Small D. Instrumental Variables Estimation With Some Invalid Instruments and its Application to Mendelian Randomization. Journal of the American Statistical Association. 2016; 111(513):132–144. https://doi.org/10.1080/01621459.2014.994705
25. Windmeijer F, Farbmacher H, Davies N, Davey Smith G. On the Use of the Lasso for Instrumental Variables Estimation with Some Invalid Instruments. Journal of the American Statistical Association. 2018; 0(0):1–12. https://doi.org/10.1080/01621459.2018.1498346
26. Cheng X, Liao Z. Select the valid and relevant moments: An information-based LASSO for GMM with many moments. Journal of Econometrics. 2015; 186(2):443–464. https://doi.org/10.1016/j.jeconom.2015.02.019
27. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society, Series B (Methodological). 1996; 58:267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

28. Goeman J, Meijer R, Chaturvedi N, Lueder M. penalized: L1 (Lasso and Fused Lasso) and L2 (Ridge) Penalized Estimation in GLMs and in the Cox Model; 2017 URL https://cran.r-project.org/web/packages/penalized/penalized.pdf.
29. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics. 2016; 32(20):3207–3209. https://doi.org/10.1093/bioinformatics/btw373 PMID: 27318201
30. Coodin S. Body mass index in persons with schizophrenia. Canadian Journal of Psychiatry. 2001; 46 (6):549–555. https://doi.org/10.1177/070674370104600610 PMID: 11526812
31. Allison DB, Fontaine KR, Heo M, Mentore JL, Cappelleri JC, Chandler LP, et al. The distribution of body mass index among individuals with and without schizophrenia. Journal of Clinical Psychiatry. 1999; 60 (4):215–220. https://doi.org/10.4088/jcp.v60n0402 PMID: 10221280
32. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015; 518(7538):197–206. https://doi.org/10.1038/nature14177 PMID: 25673413
33. Ripke S, Neale BM, Corvin A, Walters JT, Farh KH, Holmans PA, et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511(7510):421–427. https://doi.org/10.1038/nature13595
34. Hartwig FP, Bowden J, Loret de Mola C, Tovo-Rodrigues L, Davey Smith G, Horta BL. Body mass index and psychiatric disorders: a Mendelian randomization study. Scientific Reports. 2016; 6:32730. https://doi.org/10.1038/srep32730 PMID: 27601421
35. Notkola IL, Sulkava R, Pekkanen J, Erkinjuntti T, Ehnholm C, Kivinen P, et al. Serum total cholesterol, apolipoprotein E epsilon 4 allele, and Alzheimer's disease. Neuroepidemiology. 1998; 17(1):14–20. https://doi.org/10.1159/000026149 PMID: 9549720
36. Solomon A, Kareholt I, Ngandu T, Winblad B, Nissinen A, Tuomilehto J, et al. Serum cholesterol changes after midlife and late-life cognition: twenty-one-year follow-up study. Neurology. 2007; 68 (10):751–756. https://doi.org/10.1212/01.wnl.0000256368.57375.b7 PMID: 17339582
37. Shepardson NE, Shankar GM, Selkoe DJ. Cholesterol level and statin use in Alzheimer disease: I. Review of epidemiological and preclinical studies. JAMA Neurology. 2011; 68(10):1239–1244.
38. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. Nature Genetics. 2013; 45(11):1345–1352. https://doi.org/10.1038/ng.2795 PMID: 24097064
39. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nature Genetics. 2013; 45(11):1274–1283. https://doi.org/10.1038/ng.2797 PMID: 24097068
40. Benn M, Nordestgaard BG, Frikke-Schmidt R, Tybjærg-Hansen A. Low LDL cholesterol, PCSK9 and HMGCR genetic variation, and risk of Alzheimer's disease and Parkinson's disease: Mendelian randomisation study. British Medical Journal. 2017; 357. https://doi.org/10.1136/bmj.j1648 PMID: 28438747
41. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature Genetics. 2013; 45(12):1452–1458. https://doi.org/10.1038/ng.2802 PMID: 24162737
42. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I^2 statistic. International Journal of Epidemiology. 2016; 45(6):1961–1974. https://doi.org/10.1093/ije/dyw220 PMID: 27616674
43. Slob EA, Burgess S. A comparison of robust Mendelian randomization methods using summary data. bioRxiv. 2019; p. 577940.
44. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Human Molecular Genetics. 2018; 27(R2):R195–R208. https://doi.org/10.1093/hmg/ddy163 PMID: 29771313
45. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. European Journal of Epidemiology. 2017; 32(5):377–389. https://doi.org/10.1007/s10654-017-0255-x PMID: 28527048

Appendix C

Appendix to paper 1

Appendix to the published paper contained in Appendix B.

S1 Software code

We provide R code to implement the methods discussed in this paper. The associations of the genetic variants with the exposure are denoted `betaXG` with standard errors `sebetaXG`. The associations of the genetic variants with the outcome are denoted `betaYG` with standard errors `sebetaYG`. We assume that the genetic variants are uncorrelated.

Inverse-variance weighted estimate

The inverse-variance weighted (IVW) estimate can be calculated by weighted linear regression:

```
betaIVW              = summary(lm(betaYG~betaXG-1, weights=sebetaYG^-2))$coef[1]
sebetaIVW.fixed      = summary(lm(betaYG~betaXG-1, weights=sebetaYG^-2))$coef[1,2]/
                       summary(lm(betaYG~betaXG-1, weights=sebetaYG^-2))$sigma
sebetaIVW.random     = summary(lm(betaYG~betaXG-1, weights=sebetaYG^-2))$coef[1,2]/
                       min(summary(lm(betaYG~betaXG-1, weights=sebetaYG^-2))$sigma,1)
```

In the fixed-effect model, we divide the reported standard error by the estimated residual standard error to force the residual standard error to be 1. In the multiplicative random-effects model, we divide by the estimated residual standard error when the variability in the genetic associations is less than expected by chance (underdispersion). When there is evidence of heterogeneity between the causal estimates (overdispersion) the standard error is unaltered. The multiplicative random-effects model will result in a larger standard error compared to the fixed-effect model if there is heterogeneity between the causal estimates. The point estimate is unaffected by the choice of a fixed- or multiplicative random-effects model.

Alternatively, the inverse-variance weighted estimate can be calculated by meta-analysis, or via a simple formula:

```
# meta-analysis
library(meta)
betaIVW              = metagen(betaYG/betaXG, abs(sebetaYG/betaXG))$TE.fixed
sebetaIVW.fixed      = metagen(betaYG/betaXG, abs(sebetaYG/betaXG))$seTE.fixed
# simple formula
betaIVW              = sum(betaYG*betaXG*sebetaYG^-2)/sum(betaXG^2*sebetaYG^-2)
sebetaIVW.fixed      = 1/sqrt(sum(betaXG^2*sebetaYG^-2))
```

MR-Egger regression

The MR-Egger method is equivalent to the IVW method calculated using weighted regression, except that that intercept term is estimated rather than being set to zero. A test as to whether the intercept term is equal to zero is a test of directional pleiotropy and the validity of the InSIDE assumption. The genetic associations with the risk factor `betaXG` and outcome `betaYG` must be orientated with respect to the risk increasing or decreasing allele of the risk factor. A random-effects model should be used for inference as a fixed-effect model is not justifiable when the genetic variants are not all valid instruments.

```
# coding of genetic variants
betaYG = betaYG*sign(betaXG); betaXG = abs(betaXG)
# causal estimate
betaEGGER           = summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$coef[2,1]
sebetaEGGER.random  = summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$coef[2,2]/
                      min(summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$sigma, 1)
betaEGGER.lower     = betaEGGER-qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random
betaEGGER.upper     = betaEGGER+qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random
p.causal.random     = 2*(1-pt(abs(betaEGGER/sebetaEGGER.random),df=length(betaXG)-2))
# test for directional pleiotropy
interEGGER          = summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$coef[1,1]
seinterEGGER.random = summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$coef[1,2]/
                      min(summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$sigma, 1)
p.dpleio.random     = 2*(1-pt(abs(interEGGER/seinterEGGER.random),df=length(betaXG)-2))
```

In this code, we use a t-distribution with $J-2$ degrees of freedom for inference. If there is underdispersion, then the t-distribution may be overly conservative, as the t-distribution assumes that the residual standard error is estimated (in case of underdispersion, the residual standard error is set to 1). Hence, if the residual standard error is less than one, either a confidence interval using a residual standard error of 1 and a z-distribution, or else a confidence interval using the estimated residual standard error and a t-distribution may be preferred (the wider of these two intervals should be preferred – both of these will be narrower than the above confidence interval).

```
sigmaEGGER         = summary(lm(betaYG~betaXG, weights=sebetaYG^-2))$sigma
betaEGGER.lower    = ifelse(sigmaEGGER<1, min(betaEGGER-qnorm(0.975)*sebetaEGGER.random,
                      betaEGGER-qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random*sigmaEGGER),
                      betaEGGER-qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random)
betaEGGER.upper    = ifelse(sigmaEGGER<1, max(betaEGGER+qnorm(0.975)*sebetaEGGER.random,
                      betaEGGER+qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random*sigmaEGGER),
                      betaEGGER+qt(0.975,df=length(betaXG)-2)*sebetaEGGER.random)
```

Median-based method

The median-based method calculates the median (or weighted median) of the causal estimates from each candidate instrument. This code calculates the simple median and weighted median, employing bootstrapping to obtain a standard error that can be used to provide a confidence interval.

```
weighted.median <- function(betaIV.in, weights.in) {
  betaIV.order  = betaIV.in[order(betaIV.in)]
  weights.order = weights.in[order(betaIV.in)]
  weights.sum   = cumsum(weights.order)-0.5*weights.order
  weights.sum   = weights.sum/sum(weights.order)
        below   = max(which(weights.sum<0.5))
  weighted.est  = betaIV.order[below] + (betaIV.order[below+1]-betaIV.order[below])*
                  (0.5-weights.sum[below])/(weights.sum[below+1]-weights.sum[below])
  return(weighted.est) }
 #
weighted.median.boot = function(betaXG.in, betaYG.in, sebetaXG.in, sebetaYG.in, weights.in){
 # the standard error is estimated based on 1000 bootstrap samples
med = NULL
for(i in 1:1000){
 betaXG.boot = rnorm(length(betaXG.in), mean=betaXG.in, sd=sebetaXG.in)
 betaYG.boot = rnorm(length(betaYG.in), mean=betaYG.in, sd=sebetaYG.in)
 betaIV.boot = betaYG.boot/betaXG.boot
 med[i]      = weighted.median(betaIV.boot, weights.in)
 }
return(sd(med)) }
 #
```

```
betaIV             = betaYG/betaXG
weights            = rep(1, length(betaXG)) # unweighted median
betaSIMPLEMED      = weighted.median(betaIV, weights)
sebetaSIMPLEMED    = weighted.median.boot(betaXG, betaYG, sebetaXG, sebetaYG, weights)
lowerSIMPLEMED     = betaSIMPLEMED-qnorm(0.975)*sebetaSIMPLEMED
upperSIMPLEMED     = betaSIMPLEMED+qnorm(0.975)*sebetaSIMPLEMED
 #
betaIV             = betaYG/betaXG
weights            = (sebetaYG/betaXG)^-2   # weighted median using inverse-variance weights
betaWEIGHTEDMED    = weighted.median(betaIV, weights)
sebetaWEIGHTEDMED  = weighted.median.boot(betaXG, betaYG, sebetaXG, sebetaYG, weights)
lowerWEIGHTEDMED   = betaWEIGHTEDMED-qnorm(0.975)*sebetaWEIGHTEDMED
upperWEIGHTEDMED   = betaWEIGHTEDMED+qnorm(0.975)*sebetaWEIGHTEDMED
```

Robust regression

The IVW and MR-Egger methods can be performed using robust regression (in particular, MM-estimation using Tukey's bisquare objective function) rather than standard linear regression:

```
library(robustbase)
betaIVW.robust            = summary(lmrob(betaYG~betaXG-1, weights=sebetaYG^-2, k.max=500))$coef[1]
sebetaIVW.robust.fixed    = summary(lmrob(betaYG~betaXG-1, weights=sebetaYG^-2, k.max=500))$coef[1,2]/
                            summary(lmrob(betaYG~betaXG-1, weights=sebetaYG^-2, k.max=500))$sigma
sebetaIVW.robust.random   = summary(lmrob(betaYG~betaXG-1, weights=sebetaYG^-2, k.max=500))$coef[1,2]/
                            min(summary(lmrob(betaYG~betaXG-1, weights=sebetaYG^-2, k.max=500))$sigma,1)
betaEGGER.robust          = summary(lmrob(betaYG~betaXG,   weights=sebetaYG^-2, k.max=500))$coef[2]
sebetaEGGER.robust.random = summary(lmrob(betaYG~betaXG,   weights=sebetaYG^-2, k.max=500))$coef[2,2]/
                            min(summary(lmrob(betaYG~betaXG,   weights=sebetaYG^-2, k.max=500))$sigma,1)
```

The `k.max` option sets the maximum number of steps evaluated to find initial parameter values in the S-step of the algorithm. The `lmrob` function sets the tuning parameter to 1.548 to provide a high breakdown point in the S-estimation step, and as 4.685 to provide efficiency in the M-estimation step.

Penalized weights

The IVW and MR-Egger methods can be performed using penalized weights:

```
betaIVW                  = sum(betaYG*betaXG*sebetaYG^-2)/sum(betaXG^2*sebetaYG^-2)
pweights                 = pchisq(betaXG^2/sebetaYG^2*(betaYG/betaXG-betaIVW)^2, df=1, lower.tail=FALSE)
pweightsE                = pchisq(sebetaYG^-2*(betaYG - interEGGER - betaEGGER*betaXG)^2, df=1, lower.tail=FALSE)
rweights                 = sebetaYG^-2*pmin(1, pweights*100)
rweightsE                = sebetaYG^-2*pmin(1, pweightsE*100)
betaIVW.penal            = summary(lm(betaYG~betaXG-1, weights=rweights))$coef[1]
sebetaIVW.penal.fixed    = summary(lm(betaYG~betaXG-1, weights=rweights))$coef[1,2]/
                           summary(lm(betaYG~betaXG-1, weights=rweights))$sigma
sebetaIVW.penal.random   = summary(lm(betaYG~betaXG-1, weights=rweights))$coef[1,2]/
                           min(summary(lm(betaYG~betaXG-1, weights=rweights))$sigma,1)
betaEGGER.penal          = summary(lm(betaYG~betaXG,   weights=rweightsE))$coef[2]
sebetaEGGER.penal.random = summary(lm(betaYG~betaXG,   weights=rweightsE))$coef[2,2]/
                           min(summary(lm(betaYG~betaXG,   weights=rweightsE))$sigma,1)
```

Penalized weights can also be used in conjunction with robust regression.

Lasso penalization

Several packages are available for running Lasso penalization methods. We chose the *penalized* package as this gave an option for some of the coefficients in the model to be penalized (the pleiotropy intercept parameters), and others not to be penalized (the causal effect parameter):

```
library(penalized)
# dividing the association estimates by sebetaYG is equivalent to weighting by
    sebetaYG^-2
betaYGw = betaYG/sebetaYG
betaXGw = betaXG/sebetaYG
pleio   = diag(rep(1, length(betaXG)))

# values of lambda for heterogeneity stopping rule
l1grid = c(seq(from=0.1, to=5, by=0.1), seq(from=5.2, to=10, by=0.2))
l1grid_rse = NULL; l1grid_length = NULL; l1grid_beta = NULL; l1grid_se = NULL

for (i in 1:length(l1grid)) {
 l1grid_which     = which(attributes(penalized(betaYGw, pleio, betaXGw,
                    lambda1=l1grid[i], trace=FALSE))$penalized==0)
 l1grid_rse[i]    = summary(lm(betaYG[l1grid_which]~betaXG[l1grid_which]-1,
                    weights=sebetaYG[l1grid_which]^-2))$sigma
 l1grid_length[i] = length(l1grid_which)
 l1grid_beta[i]   = lm(betaYG[l1grid_which]~betaXG[l1grid_which]-1,
                    weights=sebetaYG[l1grid_which]^-2)$coef[1]
 l1grid_se[i]     = summary(lm(betaYG[l1grid_which]~betaXG[l1grid_which]-1,
                    weights=sebetaYG[l1grid_which]^-2))$coef[1,2]/
                    min(summary(lm(betaYG[l1grid_which]~betaXG[l1grid_which]-1,
                    weights=sebetaYG[l1grid_which]^-2))$sigma, 1)
 }

# heterogeneity criterion for choosing lambda
l1which_hetero = c(which(l1grid_rse[1:(length(l1grid)-1)]>1
                 &diff(l1grid_rse)>qchisq(0.95, df=1)/
                 l1grid_length[2:length(l1grid)]), length(l1grid))[1]
l1hetero_beta  = l1grid_beta[l1which_hetero]
l1hetero_se    = l1grid_se[l1which_hetero]

# cross-validation criterion for choosing lambda
l1xval_lambda  = optL1(betaYGw, pleio, betaXGw)$lambda
l1xval_which   = which(attributes(penalized(betaYGw, pleio, betaXGw,
                 lambda1=l1xval_lambda))$penalized==0)
l1xval_beta    = summary(lm(alpy[l1xval_which]~alpx[l1xval_which]-1,
                 weights=alpysd[l1xval_which]^-2))$coef[1]
l1xval_se      = summary(lm(alpy[l1xval_which]~alpx[l1xval_which]-1,
                 weights=alpysd[l1xval_which]^-2))$coef[1,2]/
                 min(summary(lm(alpy[l1xval_which]~alpx[l1xval_which]-1,
                 weights=alpysd[l1xval_which]^-2))$sigma, 1)
```

We found that our choice of values of λ $(0.1, 0.2, \ldots, 4.9, 5.0, 5.2, 5.4, \ldots, 9.8, 10.0)$ worked well in both the simulations and the applied example. However, for different sets of association estimates, a different choice of values may be preferred. Additionally, particularly with large numbers of variants, a more dense choice of values may be preferred to ensure that at most one variant is added to the analysis at each incremental step.

Regression diagnostics

Studentized residuals and Cook's distance can be calculated for the IVW method as:

```
rstudent(lm(betaYG~betaXG-1, weights=sebetaYG^-2))
cooks.distance(lm(betaYG~betaXG-1, weights=sebetaYG^-2))
```

Correlated variants

The IVW method can be performed for correlated variants using the standard linear regression command after weighting the data by the Cholesky decomposition of a weighting matrix `Omega` that accounts for the inverse-variance weights and the correlation between the genetic variants, where `rho` is the (signed) correlation matrix whose (i, j)th element is the correlation between genetic variant i and genetic variant j:

```
Omega   = sebetaYG%o%sebetaYG*rho
c_betaXG = solve(t(chol(Omega)))%*%betaXG
c_betaYG = solve(t(chol(Omega)))%*%betaYG
beta_IVWcorrel  = lm(c_betaYGc_betaXG1)$coef[1]
se_IVWcorrel.fixed  = sqrt(1/(t(betaXG)%*%solve(Omega)%*%betaXG))
se_IVWcorrel.random = sqrt(1/(t(betaXG)%*%solve(Omega)%*%betaXG))*
                      max(summary(lm(c_betaYGc_betaXG1))$sigma,1)
```

The robust and penalized methods can be implemented for correlated variants similarly by replacing the `lm` command with the appropriate function.

Genetic variants and genetic associations

We provide R code to obtain data for the applied examples from PhenoScanner, a database of genetic association estimates. This code is included in the `MendelianRandomization` package for R that is available through The Comprehensive R Archive Network (CRAN) with kind permission of James Staley.

Body mass index and schizophrenia

In total, 97 genetic variants (listed below) were used in the analysis of body mass index and schizophrenia. At the time of writing, PhenoScanner was not able to process quadrallelic genetic variants, and so it only retrieved data on 95 of the variants. It is hoped that this will be fixed in the near future.

```
library(MendelianRandomization)

bmi_snps = scan(what="character")
rs1558902 rs6567160 rs13021737 rs10938397 rs543874 rs2207139 rs11030104 rs3101336 rs7138803
 rs10182181 rs3888190 rs1516725 rs12446632 rs2287019 rs16951275 rs3817334 rs2112347 rs12566985
 rs3810291 rs7141420 rs13078960 rs10968576 rs17024393 rs657452 rs12429545 rs12286929 rs13107325
 rs11165643 rs7903146 rs10132280 rs17405819 rs6091540 rs1016287 rs4256980 rs17094222 rs12401738
 rs7599312 rs2365389 rs205262 rs2820292 rs12885454 rs9641123 rs12016871 rs16851483 rs1167827
 rs758747 rs1928295 rs9925964 rs11126666 rs2650492 rs6804842 rs12940622 rs7164727 rs11847697
 rs4740619 rs492400 rs13191362 rs3736485 rs17001654 rs11191560 rs2080454 rs7715256 rs2176040
 rs1528435 rs2075650 rs1000940 rs2033529 rs11583200 rs7239883 rs2836754 rs9400239 rs10733682
 rs11688816 rs11057405 rs9914578 rs977747 rs2121279 rs29941 rs11727676 rs3849570 rs9374842
 rs6477694 rs4787491 rs1441264 rs7899106 rs2176598 rs2245368 rs17203016 rs17724992 rs7243357
 rs16907751 rs1808579 rs13201877 rs2033732 rs9540493 rs1460676 rs6465468

bmi_obj = pheno_input(snps=bmi_snps,
 exposure = "Body mass index", pmidE = "25673413", ancestryE = "European",
 outcome = "Schizophrenia", pmidO = "25056061", ancestryO = "Mixed")

mr_ivw(bmi_obj)                  # IVW method
mr_egger(bmi_obj)                # MR-Egger method
mr_median(bmi_obj)               # Weighted median method
mr_ivw(bmi_obj, robust = TRUE)   # Robust regression method
mr_ivw(bmi_obj, penalized = TRUE) # Penalized weights method
```

LDL-cholesterol and Alzheimer's disease

In total, 75 genetic variants (listed below) were used in the analysis of LDL-cholesterol and Alzheimer's disease. Data for all these variants are available in PhenoScanner.

```
ldl_snps = scan(what="character")
rs10903129 rs4587594 rs6603981 rs646776 rs1010167 rs267733 rs2642438 rs903319 rs2587534
 rs1367117 rs515135 rs6544713 rs4148218 rs2710642 rs17508045 rs2030746 rs16831243 rs1250229
 rs11563251 rs9875338 rs7640978 rs17345563 rs7703051 rs4530754 rs6882076 rs2294261 rs1800562
 rs2247056 rs868943 rs2297374 rs1564348 rs12670798 rs4722551 rs2073547 rs217386 rs4240624
 rs10102164 rs2326077 rs2737252 rs2980885 rs2954022 rs7832643 rs3780181 rs1883025 rs8176720
 rs579459 rs2255141 rs10832962 rs174532 rs1535 rs10790162 rs11220462 rs653178 rs6489818 rs1169288
 rs4942486 rs8017377 rs9989419 rs2288002 rs2000999 rs314253 rs7225700 rs6511720 rs688 rs10401969
 rs6859 rs7254892 rs492602 rs364585 rs2328223 rs7264396 rs6016381 rs6065311 rs1800961 rs5763662

ldl_obj = pheno_input(snps=ldl_snps,
 exposure = "Low density lipoprotein", pmidE = "24097068", ancestryE = "European",
 outcome = "Alzheimers disease", pmidO = "24162737", ancestryO = "European")

mr_ivw(ldl_obj)                  # IVW method (and so on)
```

The genetic association estimates can also be obtained from the web-based version of PhenoScanner at `http://www.phenoscanner.medschl.cam.ac.uk/`.

S2 Supplementary tables from the simulation study

Mean values for the R^2, F-statistic and I^2 statistic

Table A contains the mean R^2 (%), F-statistic and I^2 (%) from the simulation study for all scenarios considered.

Table A. Mean values of the R^2 (%), F-statistic and I^2 (%) for Scenarios 1-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instrumental variables (IV).

	No invalid IVs			1 invalid IV			3 invalid IVs			6 invalid IVs		
	R^2	F	I^2	R^2	F	I^2	R^2	F	I^2	R^2	F	I^2
Null causal effect: $\theta = 0$												
Scenario 1	3.0	20.8	39.6	-	-	-	-	-	-	-	-	
Scenario 2	-	-	-	3.0	20.8	39.6	3.0	20.8	39.3	3.0	20.8	39.5
Scenario 3	-	-	-	3.0	20.8	39.7	3.0	20.8	39.5	3.0	20.8	39.2
Scenario 4	-	-	-	3.4	23.6	56.5	4.2	29.3	70.7	5.4	37.7	77.5
Positive causal effect: $\theta = 0.3$												
Scenario 1	3.0	20.8	39.3	-	-	-	-	-	-	-	-	
Scenario 2	-	-	-	3.0	20.8	39.1	3.0	20.8	39.4	3.0	20.8	39.6
Scenario 3	-	-	-	3.0	20.8	39.9	3.0	20.8	39.7	3.0	20.8	39.6
Scenario 4	-	-	-	3.4	23.6	56.4	4.2	29.3	70.8	5.4	37.7	77.4

Number of robust regression analyses without a standard error

The number of robust regressions that did not report a standard error in the simulations are presented in Table B. The proportion of simulations was less than 1.2% across the different scenarios for the IVW model. Apart from the calculation of the mean standard error, the simulations that did not report a standard error were included in the results, and the power calculations treated the standard error as infinite.

Table B. Number of the 10 000 simulations that failed to report a standard error using robust regression (without and with penalized weights) with the inverse-variance weighted (IVW) and MR-Egger methods, for Scenarios 1-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instrumental variables.

	IVW								MR-Egger							
	Robust				Robust, penalized				Robust				Robust, penalized			
No. invalid:	0	1	3	6	0	1	3	6	0	1	3	6	0	1	3	6
Null causal effect: $\theta = 0$																
Scenario 1	0	-	-	-	0	-	-	-	16	-	-	-	16	-	-	-
Scenario 2	-	1	2	5	-	3	9	120	-	24	72	78	-	45	98	258
Scenario 3	-	2	1	4	-	3	10	51	-	32	69	32	-	30	70	139
Scenario 4	-	3	84	11	-	5	6	22	-	144	100	5	-	124	76	9
Positive causal effect: $\theta = 0.3$																
Scenario 1	4	-	-	-	3	-	-	-	13	-	-	-	13	-	-	-
Scenario 2	-	0	0	1	-	1	3	54	-	20	55	47	-	24	71	211
Scenario 3	-	0	0	3	-	3	2	22	-	24	72	19	-	37	62	73
Scenario 4	-	2	30	4	-	0	9	9	-	151	91	10	-	122	81	15

Abbreviations: IVW, inverse-variance weighted; No., number.

MR-Egger intercept test

Table C contains information on the power (at the 5% significance level) of the intercept test in the MR-Egger method for detecting directional pleiotropy and/or violation of the InSIDE assumption for all scenarios.

Table C. Power (%) of the intercept test in the MR-Egger method for detecting directional pleiotropy and/or violation of the InSIDE assumption for Scenarios 1-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instrumental variables (IV).

	Null causal effect				Positive causal effect			
No. invalid:	0	1	3	6	0	1	3	6
Scenario 1	3.7	-	-	-	8.7	-	-	
Scenario 2	-	7.2	7.5	7.0	-	9.4	8.5	7.8
Scenario 3	-	7.2	8.7	13.1	-	11.2	13.8	19.1
Scenario 4	-	22.8	49.9	55.9	-	8.6	26.2	32.0

Results from applying the robust methods to the MR-Egger method

Table D contains the results from the simulation study when the MR-Egger model was applied to the simulated data with: 1) robust regression (R); 2) penalized weights (P); and 3) robust regression and penalized weights (R and P).

Table D. Mean (standard error) estimates and power from the MR-Egger method with: robust regression (R); penalized weights (P); and robust regression and penalized weights (R and P) for Scenarios 1-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instrumental variables.

	No invalid IVs		1 invalid IV		3 invalid IVs		6 invalid IVs	
	Mean (mean SE)	Power, %	Mean (mean SE)	Power, %	Mean (mean SE)	Power, %	Mean (mean SE)	Power, %
Null causal effect: $\theta = 0$								
Scenario 1. No pleiotropy, InSIDE satisfied								
R	0.000 (0.231)	8.2	-	-	-	-	-	-
P	-0.001 (0.216)	4.3	-	-	-	-	-	-
R and P	0.000 (0.230)	8.3	-	-	-	-	-	-
Scenario 2. Balanced pleiotropy, InSIDE satisfied								
R	-	-	-0.006 (0.245)	9.7	-0.002 (0.375)	9.6	-0.007 (0.671)	10.8
P	-	-	-0.006 (0.208)	9.9	-0.003 (0.231)	16.9	-0.009 (0.274)	31.5
R and P	-	-	-0.007 (0.254)	9.2	-0.001 (0.333)	10.7	-0.008 (0.505)	20.3
Scenario 3. Directional pleiotropy, InSIDE satisfied								
R	-	-	-0.003 (0.246)	9.9	0.001 (0.376)	9.5	-0.004 (0.564)	13.8
P	-	-	-0.004 (0.208)	10.1	0.001 (0.249)	18.7	-0.009 (0.343)	37.9
R and P	-	-	-0.004 (0.256)	9.4	0.001 (0.309)	12.0	-0.005 (0.419)	32.4
Scenario 4. Directional pleiotropy, InSIDE violated								
R	-	-	0.171 (0.291)	18.2	0.493 (0.234)	65.1	0.649 (0.158)	95.6
P	-	-	0.241 (0.196)	33.0	0.527 (0.178)	81.2	0.651 (0.159)	97.7
R and P	-	-	0.173 (0.272)	18.5	0.490 (0.215)	68.2	0.652 (0.148)	96.8
Positive causal effect: $\theta = 0.3$								
Scenario 1. No pleiotropy, InSIDE satisfied								
R	0.144 (0.273)	13.1	-	-	-	-	-	-
P	0.143 (0.258)	7.9	-	-	-	-	-	-
R and P	0.144 (0.271)	13.3	-	-	-	-	-	-
Scenario 2. Balanced pleiotropy, InSIDE satisfied								
R	-	-	0.140 (0.295)	13.1	0.139 (0.430)	11.6	0.124 (0.665)	11.9
P	-	-	0.139 (0.255)	13.5	0.140 (0.282)	19.4	0.130 (0.331)	30.2
R and P	-	-	0.140 (0.297)	12.9	0.140 (0.363)	14.6	0.133 (0.508)	22.4
Scenario 3. Directional pleiotropy, InSIDE satisfied								
R	-	-	0.141 (0.295)	13.2	0.135 (0.433)	11.8	0.136 (0.563)	14.7
P	-	-	0.140 (0.252)	13.2	0.137 (0.302)	19.9	0.137 (0.392)	30.7
R and P	-	-	0.140 (0.292)	13.3	0.135 (0.352)	15.9	0.138 (0.433)	31.0
Scenario 4. Directional pleiotropy, InSIDE violated								
R	-	-	0.338 (0.340)	25.5	0.719 (0.274)	75.6	0.893 (0.190)	97.7
P	-	-	0.418 (0.233)	48.9	0.754 (0.210)	91.2	0.895 (0.188)	99.4
R and P	-	-	0.340 (0.319)	27.1	0.716 (0.249)	78.6	0.897 (0.179)	98.2

Abbreviations: IV, instrumental variable; SE, standard error; InSIDE, instrument strength independent of direct effect; R, robust regression; P, penalized weights.

Results from the one-sample setting

Results from the simulation study when the data were generated from one sample are contained in Table E (null causal effect $\theta = 0$) and Table F (positive causal effect $\theta = 0.3$). Estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (R); 3) penalized weights (P); and 4) robust regression and penalized weights (R and P), and the Lasso penalization (LP) method with the heterogeneity stopping rule for Scenarios 1-4 are displayed in the Tables E and F.

Table E. Mean (standard error) and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (R); 3) penalized weights (P); and 4) robust regression and penalized weights (R and P) for Scenarios 1-4 with a null causal effect ($\theta = 0$) by the number of invalid instrumental variables for one-sample Mendelian randomization. Results from the Lasso penalization (LP) method with the heterogeneity stopping rule are also provided.

	No invalid IVs		1 invalid IV		3 invalid IVs		6 invalid IVs	
	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %
Null causal effect: $\boldsymbol{\theta = 0}$								
Scenario 1. No pleiotropy, InSIDE satisfied								
IVW	0.021 (0.061)	5.6	-	-	-	-	-	-
R	0.021 (0.065)	6.8	-	-	-	-	-	-
P	0.019 (0.060)	6.0	-	-	-	-	-	-
R and P	0.020 (0.063)	7.2	-	-	-	-	-	-
LP	0.021 (0.060)	6.2	-	-	-	-	-	-
Scenario 2. Balanced pleiotropy, InSIDE satisfied								
IVW	-	-	0.020 (0.088)	6.4	0.020 (0.132)	7.1	0.024 (0.180)	7.3
R	-	-	0.021 (0.068)	7.8	0.020 (0.096)	6.8	0.023 (0.195)	5.9
P	-	-	0.018 (0.062)	7.1	0.015 (0.066)	9.7	0.008 (0.075)	19.7
R and P	-	-	0.019 (0.070)	6.8	0.017 (0.092)	6.3	0.010 (0.156)	7.7
LP	-	-	0.020 (0.063)	7.1	0.020 (0.070)	9.2	0.019 (0.088)	16.9
Scenario 3. Directional pleiotropy, InSIDE satisfied								
IVW	-	-	0.086 (0.088)	9.3	0.216 (0.123)	36.2	0.409 (0.150)	92.5
R	-	-	0.032 (0.067)	8.8	0.088 (0.109)	11.2	0.357 (0.222)	44.9
P	-	-	0.025 (0.062)	7.0	0.046 (0.067)	13.6	0.132 (0.081)	40.4
R and P	-	-	0.025 (0.070)	7.5	0.040 (0.088)	10.8	0.103 (0.125)	21.5
LP	-	-	0.027 (0.063)	7.2	0.049 (0.071)	12.6	0.173 (0.096)	42.8
Scenario 4. Directional pleiotropy, InSIDE violated								
IVW	-	-	0.096 (0.068)	27.5	0.202 (0.072)	86.6	0.303 (0.067)	100.0
R	-	-	0.053 (0.081)	10.1	0.163 (0.119)	38.4	0.302 (0.072)	98.4
P	-	-	0.040 (0.061)	14.2	0.095 (0.062)	41.4	0.237 (0.061)	89.6
R and P	-	-	0.038 (0.069)	11.4	0.089 (0.079)	30.7	0.236 (0.071)	83.9
LP	-	-	0.048 (0.062)	17.2	0.138 (0.066)	58.8	0.300 (0.064)	99.1

Abbreviations: IV, instrumental variable; SE, standard error; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted; R, robust regression; P, penalized weights; LP, lasso penalization.

Table F. Mean (standard error) and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (R); 3) penalized weights (P); and 4) robust regression and penalized weights (R and P) for Scenarios 1-4 with a positive causal effect ($\theta = 0.3$) by the number of invalid instrumental variables for one-sample Mendelian randomization. Results from the Lasso penalization (LP) method with the heterogeneity stopping rule are also provided.

	No invalid IVs		1 invalid IV		3 invalid IVs		6 invalid IVs	
	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %
Positive causal effect: $\boldsymbol{\theta = 0.3}$								
Scenario 1. No pleiotropy, InSIDE satisfied								
IVW	0.321 (0.068)	99.7	-	-	-	-	-	-
R	0.321 (0.073)	97.7	-	-	-	-	-	-
P	0.321 (0.068)	99.7	-	-	-	-	-	-
R and P	0.321 (0.072)	97.9	-	-	-	-	-	-
LP	0.321 (0.068)	99.7	-	-	-	-	-	-
Scenario 2. Balanced pleiotropy, InSIDE satisfied								
IVW	-	-	0.322 (0.090)	91.2	0.322 (0.133)	66.2	0.323 (0.180)	43.9
R	-	-	0.322 (0.073)	97.2	0.321 (0.098)	86.1	0.323 (0.193)	43.6
P	-	-	0.320 (0.070)	99.1	0.316 (0.076)	96.4	0.308 (0.087)	85.4
R and P	-	-	0.321 (0.078)	96.3	0.317 (0.096)	88.4	0.310 (0.138)	67.3
LP	-	-	0.322 (0.071)	99.2	0.320 (0.079)	95.6	0.320 (0.099)	81.4
Scenario 3. Directional pleiotropy, InSIDE satisfied								
IVW	-	-	0.386 (0.088)	99.7	0.517 (0.124)	100.0	0.710 (0.150)	100.0
R	-	-	0.332 (0.073)	97.9	0.390 (0.111)	93.3	0.655 (0.226)	86.8
P	-	-	0.330 (0.070)	99.7	0.362 (0.076)	99.5	0.465 (0.093)	99.5
R and P	-	-	0.328 (0.078)	96.0	0.351 (0.093)	92.6	0.434 (0.125)	89.1
LP	-	-	0.331 (0.071)	99.6	0.363 (0.079)	99.4	0.508 (0.108)	99.2
Scenario 4. Directional pleiotropy, InSIDE violated								
IVW	-	-	0.396 (0.070)	100.0	0.502 (0.073)	100.0	0.601 (0.067)	100.0
R	-	-	0.352 (0.088)	95.5	0.463 (0.125)	89.7	0.601 (0.073)	99.7
P	-	-	0.349 (0.068)	99.7	0.424 (0.068)	99.7	0.561 (0.064)	99.9
R and P	-	-	0.343 (0.078)	97.0	0.414 (0.090)	95.7	0.562 (0.071)	98.8
LP	-	-	0.357 (0.068)	99.7	0.463 (0.070)	99.8	0.600 (0.066)	100.0

Abbreviations: IV, instrumental variable; SE, standard error; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted; R, robust regression; P, penalized weights; LP, lasso penalization.

Mean values for the R^2, F-statistic and I^2 statistic for 100 genetic variants

Table G contains the mean R^2 (%), F-statistic and I^2 (%) when the simulation study was re-performed for 100 genetic variants for Scenarios 2-4.

Table G. Mean values of the R^2 (%), F-statistic and I^2 (%) for Scenarios 2-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instrumental variables (IV) when the simulation study was re-perfomed for 100 genetic variants.

	5 invalid IV			15 invalid IVs			30 invalid IVs		
	R^2	F	I^2	R^2	F	I^2	R^2	F	I^2
Null causal effect: $\boldsymbol{\theta = 0}$									
Scenario 2	4.0	4.2	3.0	4.0	4.2	3.3	4.0	4.2	2.9
Scenario 3	4.0	4.2	3.1	4.0	4.2	3.0	4.0	4.2	3.0
Scenario 4	5.2	5.4	32.9	7.3	7.8	58.3	10.3	11.4	69.5
Positive causal effect: $\boldsymbol{\theta = 0.3}$									
Scenario 2	4.0	4.2	3.2	4.0	4.2	3.1	4.1	4.2	3.1
Scenario 3	4.0	4.2	3.3	4.1	4.2	3.1	4.0	4.2	2.9
Scenario 4	5.2	5.4	33.3	7.3	7.8	58.2	10.3	11.4	69.4

Appendix D

Paper 2: Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy

Published paper based on the work in Chapter 4.

Received: 20 October 2016 | Revised: 15 August 2017 | Accepted: 23 August 2017

DOI: 10.1002/sim.7492

RESEARCH ARTICLE

WILEY Statistics in Medicine

Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy

Jessica M. B. Rees[1] | Angela M. Wood[1] | Stephen Burgess[1,2]

[1]Cardiovascular Epidemiology Unit, University of Cambridge, Cambridge, UK

[2]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

Correspondence
Stephen Burgess, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.
Email: sb452@medschl.cam.ac.uk

Funding information
British Heart Foundation, Grant/Award Number: FS/14/59/31282; Wellcome Trust and the Royal Society, Grant/Award Number: 204623/Z/16/Z

Methods have been developed for Mendelian randomization that can obtain consistent causal estimates while relaxing the instrumental variable assumptions. These include multivariable Mendelian randomization, in which a genetic variant may be associated with multiple risk factors so long as any association with the outcome is via the measured risk factors (measured pleiotropy), and the MR-Egger (Mendelian randomization-Egger) method, in which a genetic variant may be directly associated with the outcome not via the risk factor of interest, so long as the direct effects of the variants on the outcome are uncorrelated with their associations with the risk factor (unmeasured pleiotropy). In this paper, we extend the MR-Egger method to a multivariable setting to correct for both measured and unmeasured pleiotropy. We show, through theoretical arguments and a simulation study, that the multivariable MR-Egger method has advantages over its univariable counterpart in terms of plausibility of the assumption needed for consistent causal estimation and power to detect a causal effect when this assumption is satisfied. The methods are compared in an applied analysis to investigate the causal effect of high-density lipoprotein cholesterol on coronary heart disease risk. The multivariable MR-Egger method will be useful to analyse high-dimensional data in situations where the risk factors are highly related and it is difficult to find genetic variants specifically associated with the risk factor of interest (multivariable by design), and as a sensitivity analysis when the genetic variants are known to have pleiotropic effects on measured risk factors.

KEYWORDS
invalid instruments, Mendelian randomization, MR-Egger, multivariable, pleiotropy

1 | INTRODUCTION

Mendelian randomization (MR) uses genetic variants as instrumental variables to estimate the causal effect of a risk factor on an outcome using observational data.[1,2] Increases in the scale of genome-wide association studies have led to large numbers of genetic variants that are associated with candidate risk factors being discovered.[3] If the variants

explain additional variability in the risk factor then using multiple variants in a MR analysis will increase power to detect a causal effect.[4,5] A pleiotropic genetic variant is associated with multiple risk factors; such a variant is not a valid instrumental variable and its inclusion in an (univariable) MR analysis may result in biased causal estimates and inappropriate inferences.[6] As more variants are used in an MR analysis, the chance of including a pleiotropic variant increases.

For some sets of risk factors, including lipid fractions, several risk factors have common genetic predictors. Although such genetic variants are pleiotropic, they can be used to estimate causal effects in a multivariable MR framework.[7] In multivariable MR, the instrumental variable assumptions are extended to allow a genetic variant to be associated with multiple risk factors, provided all associated risk factors are included in the analysis. Alternatively, when genetic variants are suspected to violate the instrumental variable assumptions through unknown pleiotropic pathways, methods have been developed to estimate consistent causal effects under weaker assumptions. These include the weighted median and MR-Egger methods.[8,9] The extension of MR-Egger to a multivariable setting has been implemented by Helgadottir et al as part of a sensitivity analysis in their applied work investigating the effect of lipid fractions on coronary heart disease (CHD) risk.[10] However, there remains several methodological issues relating to the implementation of the method, and the assumptions required.

In this paper, we expand univariable MR-Egger to the multivariable setting. In Section 2, we introduce the conventional and MR-Egger methods in both univariable and multivariable contexts. We provide an example analysis using published data on lipid fractions and CHD risk (Section 3), and compare results from the different MR methods in a simulation study (Section 4). Finally (Section 5), we discuss the results of the paper and the implications for applied practice. Software code for implementing all of the methods used in this paper is provided in the Web Appendix.

2 | METHODS

Initially, we consider the causal effect of a risk factor X on an outcome Y using genetic variants G_j ($j = 1, \dots, J$) that are assumed to be uncorrelated (not in linkage disequilibrium). Then, we expand to consider multiple risk factors $X_1, X_2, \dots, X_K$. Increasingly, MR investigations are implemented using summarized data from consortia to leverage their large sample sizes, thereby improving the precision of causal estimates.[11] We therefore assume that summarized data are available on the associations of each genetic variant with the risk factor (or with each risk factor for the multivariable setting) and with the outcome: the beta-coefficients ($\hat{\beta}_{X_j}, \hat{\beta}_{Y_j}$) and their standard errors ($\text{se}(\hat{\beta}_{X_j}), \text{se}(\hat{\beta}_{Y_j})$) from univariable regression on each variant G_j in turn. We additionally assume that the associations of genetic variants with the risk factor and the outcome, and the causal effect of the risk factor on the outcome, are linear and homogeneous across the population; these assumptions are discussed in detail elsewhere.[12] To distinguish between the parameters from the different methods considered, we use the following subscript notation: UI ("univariable inverse variance weighted (IVW)"); UE ("univariable MR-Egger"); MI ("multivariable IVW"); and ME ("multivariable MR-Egger").

2.1 | Univariable Mendelian randomization

In a univariable MR analysis, each genetic variant must satisfy the following criteria to be a valid instrumental variable (IV):

- IV1: The variant is associated with the risk factor X,
- IV2: The variant is independent of all confounders U of the risk factor-outcome association, and
- IV3: The variant is independent of the outcome Y conditional on the risk factor X and confounders U.[13]

These assumptions imply that the genetic variant should not have an effect on the outcome except via the risk factor. Under linearity assumptions, the association between the genetic variant and the outcome can be decomposed into an indirect effect via the risk factor and a direct effect:

$$\beta_{Y_j} = \alpha_j + \theta\beta_{X_j}, \tag{1}$$

where θ is the causal effect of the risk factor on the outcome. Genetic variant j is pleiotropic if $\alpha_j \neq 0$, and α_j is the direct effect of the genetic variant on the outcome. Figure 1 contains a direct effect α_j via an independent pathway, which violates the IV3 assumption.

With a single genetic variant, G_1 say, the causal estimate is $\hat{\beta}_{Y_1}/\hat{\beta}_{X_1}$.[14] This is a consistent estimate of the causal effect θ when $\alpha_1 = 0$. With multiple genetic variants, the inverse-variance weighted (IVW) estimate is the weighted average of

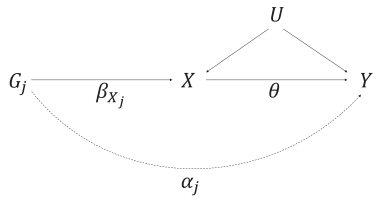


FIGURE 1 Causal directed acyclic graph illustrating univariable Mendelian randomization assumptions with potential violation of IV3 by a pleiotropic effect indicated by a dotted line. The genetic effect of G_j on X is β_{X_j}, the direct (pleiotropic) effect of G_j on Y via an independent pathway is α_j (representing the potential violation of the IV3 assumption), and the causal effect of the risk factor X on the outcome Y is θ. U represents the set of variables that confound the association between X and Y

these causal estimates,[15] using the inverse of their approximate variances $\text{se}(\hat{\beta}_{Y_j})^2/\hat{\beta}^2_{X_j}$ as weights:

$$\hat{\theta}_{UI} = \frac{\sum_j \hat{\beta}_{Y_j}\hat{\beta}_{X_j}\text{se}(\hat{\beta}_{Y_j})^{-2}}{\sum_j \hat{\beta}^2_{X_j}\text{se}(\hat{\beta}_{Y_j})^{-2}}. \tag{2}$$

This estimate can also be obtained from individual-level data using the 2-stage least squares method.[16] Alternatively, the causal effect of the risk factor on the outcome can be estimated using a weighted linear regression of the genetic association estimates,[17] with the intercept set to zero:

$$\hat{\beta}_{Y_j} = \theta_{UI}\hat{\beta}_{X_j} + \epsilon_{UI_j}, \quad \text{weights} = \text{se}(\hat{\beta}_{Y_j})^{-2}. \tag{3}$$

The above weighted regression model, where the residual standard error is set to one, is equivalent to performing a fixed-effect meta-analysis of the variant-specific causal estimates.[18] Under a multiplicative random effects model, the residual standard error can be greater than one, allowing for heterogeneity in the causal estimates. The point estimate from the fixed and random effect models will be the same, but the standard error of the causal effect from the multiplicative random effects model will be larger if there is heterogeneity between the causal estimates. Throughout this paper, we apply a multiplicative random effects model to all the analyses.

The MR-Egger estimate is obtained using the same regression model as Equation 2, but allowing the intercept to be estimated[9]:

$$\hat{\beta}_{Y_j} = \theta_{0UE} + \theta_{UE}\hat{\beta}_{X_j} + \epsilon_{UE_j}, \quad \text{weights} = \text{se}(\hat{\beta}_{Y_j})^{-2}. \tag{4}$$

If the genetic variants are not pleiotropic, then the intercept term should tend to zero as the sample size increases, and the MR-Egger estimate ($\hat{\theta}_{UE}$) and the IVW estimate ($\hat{\theta}_{UI}$) are both consistent estimates of the causal effect. Additionally, if the genetic variants are pleiotropic but the direct effects $\boldsymbol{\alpha}$ (bold symbols represent vectors across the j genetic variants) are independent of the associations of the variants with the risk factor $\boldsymbol{\beta_X}$ (known as the InSIDE assumption—Instrument Strength Independent of Direct Effect), then the MR-Egger estimate will be a consistent estimate of θ.[9,19]

Under the InSIDE assumption, the intercept term $\hat{\theta}_{0UE}$ can be interpreted as an estimate of the average direct effect of the genetic variants.[8] If the average direct effect is zero (referred to as "balanced pleiotropy"), and the InSIDE assumption is satisfied, the intercept term should tend to zero as the sample size increases, and the MR-Egger estimate ($\hat{\theta}_{UE}$) and the IVW estimate ($\hat{\theta}_{UI}$) are both consistent estimates of the causal effect. If the intercept term differs from zero, then either the InSIDE assumption is violated or the average direct effect differs from zero (referred to as "directional pleiotropy"); this is a test of the validity of the instrumental variable assumptions (the MR-Egger intercept test).

2.2 | Multivariable Mendelian randomization

In a multivariable MR analysis, each genetic variant must satisfy the following criteria:

- IV1(M): The variant is associated with at least one of the risk factors X_k,
- IV2(M): The variant is independent of all confounders U of each of the risk factor-outcome associations, and
- IV3(M): The variant is independent of the outcome Y conditional on the risk factors X_k and confounders U.[7]

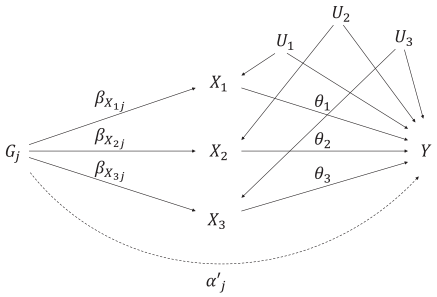


FIGURE 2 Causal directed acyclic graph illustrating multivariable Mendelian randomization assumptions for a set of genetic variants G_j, 3 risk factors X_1, X_2, and X_3, and outcome Y. The genetic effect of G_j on X_k is $\beta_{X_{kj}}$, the direct (pleiotropic) effect of G_j on Y is α'_j, and the causal effect of the risk factor X_k on the outcome Y is θ_k. U_k represents the set of variables that confound the associations between X_k and Y

Now, the association of the genetic variants with the outcome can be decomposed into indirect effects via each of the risk factors and a residual direct effect α'_j. Assuming there are 3 risk factors and all relationships are linear:

$$\beta_{Y_j} = \alpha'_j + \theta_1\beta_{X_{1j}} + \theta_2\beta_{X_{2j}} + \theta_3\beta_{X_{3j}}, \tag{5}$$

where θ_k is the causal effect of the risk factor k on the outcome (Figure 2). We assume that the risk factors do not have causal effects on each other; we later relax this assumption and allow for causal effects between the risk factors.

As in the univariable setting, causal estimates of the effect of each risk factor on the outcome can be obtained from individual-level data using the 2-stage least squares method.[7] The same estimates can also be obtained using multivariable weighted linear regression of the genetic association estimates, with the intercept set to zero (referred to as the multivariable IVW method)[20]:

$$\hat{\beta}_{Y_j} = \theta_{1MI}\hat{\beta}_{X_{1j}} + \theta_{2MI}\hat{\beta}_{X_{2j}} + \theta_{3MI}\hat{\beta}_{X_{3j}} + \epsilon_{MI_j}, \quad \text{weights} = \text{se}(\hat{\beta}_{Y_j})^{-2}. \tag{6}$$

We propose the natural extension to multivariable MR-Egger using the same regression model but allowing the intercept to be estimated:

$$\hat{\beta}_{Y_j} = \theta_{0ME} + \theta_{1ME}\hat{\beta}_{X_{1j}} + \theta_{2ME}\hat{\beta}_{X_{2j}} + \theta_{3ME}\hat{\beta}_{X_{3j}} + \epsilon_{ME_j}, \quad \text{weights} = \text{se}(\hat{\beta}_{Y_j})^{-2}. \tag{7}$$

2.3 | Assumptions for multivariable MR-Egger

We assume that the causal effect of risk factor 1 (θ_1) is of interest and provide the assumptions necessary for the MR-Egger estimate of θ_1 to be consistent. If all of the causal effects are to be interpreted, then these assumptions must apply for each risk factor.

If the $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ parameters are independent of the $\boldsymbol{\beta}_{\boldsymbol{X}_k}$ parameters for all $k = 2, 3, \ldots, K$, then the InSIDE assumption for multivariable MR-Egger is satisfied if the direct effects of the genetic variants $\boldsymbol{\alpha'}$ are independent of $\boldsymbol{\beta}_{\boldsymbol{X}_1}$. More formally, we require:

$$\boldsymbol{\beta}_{\boldsymbol{X}_1} \perp\!\!\!\perp \boldsymbol{\alpha'}, \quad \text{if } \boldsymbol{\beta}_{\boldsymbol{X}_1} \perp\!\!\!\perp \boldsymbol{\beta}_{\boldsymbol{X}_2}, \ldots, \boldsymbol{\beta}_{\boldsymbol{X}_K}, \tag{8}$$

for the estimate of θ_1 from multivariable MR-Egger to be consistent. If the InSIDE assumption is satisfied, then the weighted covariance of $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ and $\boldsymbol{\alpha'}$ ($\text{cov}_w(\boldsymbol{\alpha'}, \boldsymbol{\beta}_{\boldsymbol{X}_1})$) will tend to zero as the number of genetic variants J tends to infinity. The estimate of θ_1 from multivariable MR-Egger when the $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ parameters are independent of $\boldsymbol{\beta}_{\boldsymbol{X}_k}$ for all $k = 2, 3, \ldots, K$ is

$$\hat{\theta}_{1ME} = \frac{\text{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{Y}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X}_1})}{\text{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X}_1})} \xrightarrow{N\to\infty} \frac{\text{cov}_w(\boldsymbol{\beta}_{\boldsymbol{Y}}, \boldsymbol{\beta}_{\boldsymbol{X}_1})}{\text{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}_1})} = \theta_1 + \frac{\text{cov}_w(\boldsymbol{\alpha'}, \boldsymbol{\beta}_{\boldsymbol{X}_1})}{\text{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}_1})}, \tag{9}$$

which is equal to θ_1 if the InSIDE assumption is satisfied, where cov_w and var_w represent the weighted covariance and weighted variance using the inverse-variance weights $\text{se}(\hat{\beta}_{Yj})^{-2}$:

$$\begin{aligned}
\mathrm{cov}_w(\boldsymbol{\alpha'}, \boldsymbol{\beta}_{\boldsymbol{X}_1}) &= \frac{\sum_j (\alpha'_j - \bar{\alpha'}_w)(\beta_{X_{1j}} - \bar{\beta}_{X_1 w})\mathrm{se}(\hat{\beta}_{Yj})^{-2}}{\sum_j \mathrm{se}(\hat{\beta}_{Yj})^{-2}} \\
\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}_1}) &= \frac{\sum_j (\beta_{X_{1j}} - \bar{\beta}_{X_1 w})^2 \mathrm{se}(\hat{\beta}_{Yj})^{-2}}{\sum_j \mathrm{se}(\hat{\beta}_{Yj})^{-2}} \\
\bar{\alpha'}_w &= \frac{\sum_j \alpha'_j \mathrm{se}(\hat{\beta}_{Yj})^{-2}}{\sum_j \mathrm{se}(\hat{\beta}_{Yj})^{-2}} \\
\bar{\beta}_{X_1 w} &= \frac{\sum_j \beta_{X_{1j}} \mathrm{se}(\hat{\beta}_{Yj})^{-2}}{\sum_j \mathrm{se}(\hat{\beta}_{Yj})^{-2}}.
\end{aligned} \tag{10}$$

If the $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ parameters are correlated with at least one of the sets of $\boldsymbol{\beta}_{\boldsymbol{X}_k}$ parameters ($k = 2, 3, \ldots, K$), then the InSIDE assumption is required to hold for $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ and for all of the $\boldsymbol{\beta}_{\boldsymbol{X}_k}$ parameters that are correlated with $\boldsymbol{\beta}_{\boldsymbol{X}_1}$. More formally, we require:

$$\boldsymbol{\beta}_{\boldsymbol{X}_k} \perp\!\!\!\perp \boldsymbol{\alpha'}, \quad \text{for all } \boldsymbol{\beta}_{\boldsymbol{X}_k} \text{ correlated with } \boldsymbol{\beta}_{\boldsymbol{X}_1} (\text{including } \boldsymbol{\beta}_{\boldsymbol{X}_1} \text{itself}). \tag{11}$$

For example, if $k = 2$, and $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ is correlated with $\boldsymbol{\beta}_{\boldsymbol{X}_2}$, we require both of the weighted covariances of $\boldsymbol{\alpha'}$ with $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ and $\boldsymbol{\beta}_{\boldsymbol{X}_2}$ to be zero to produce a consistent estimate of θ_1. The estimate of θ_1 from multivariable MR-Egger with 2 risk factors where $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ and $\boldsymbol{\beta}_{\boldsymbol{X}_2}$ are correlated is

$$\begin{aligned}
\hat{\theta}_{1ME} &= \frac{\mathrm{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{Y}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X}_1})\mathrm{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X}_2}) - \mathrm{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{Y}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X}_2})\mathrm{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X}_1}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X}_2})}{\mathrm{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X}_1})\mathrm{var}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X}_2}) - \mathrm{cov}_w(\hat{\boldsymbol{\beta}}_{\boldsymbol{X}_1}, \hat{\boldsymbol{\beta}}_{\boldsymbol{X}_2})^2} \\
&\xrightarrow{N\to\infty} \frac{\mathrm{cov}_w(\boldsymbol{\beta}_{\boldsymbol{Y}}, \boldsymbol{\beta}_{\boldsymbol{X}_1})\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}_2}) - \mathrm{cov}_w(\boldsymbol{\beta}_{\boldsymbol{Y}}, \boldsymbol{\beta}_{\boldsymbol{X}_2})\mathrm{cov}_w(\boldsymbol{\beta}_{\boldsymbol{X}_1}, \boldsymbol{\beta}_{\boldsymbol{X}_2})}{\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}_1})\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}_2}) - \mathrm{cov}_w(\boldsymbol{\beta}_{\boldsymbol{X}_1}, \boldsymbol{\beta}_{\boldsymbol{X}_2})^2} \\
&= \theta_1 + \frac{\mathrm{cov}_w(\boldsymbol{\alpha'}, \boldsymbol{\beta}_{\boldsymbol{X}_1})\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}_2}) - \mathrm{cov}_w(\boldsymbol{\alpha'}, \boldsymbol{\beta}_{\boldsymbol{X}_2})\mathrm{cov}_w(\boldsymbol{\beta}_{\boldsymbol{X}_1}, \boldsymbol{\beta}_{\boldsymbol{X}_2})}{\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}_1})\mathrm{var}_w(\boldsymbol{\beta}_{\boldsymbol{X}_2}) - \mathrm{cov}_w(\boldsymbol{\beta}_{\boldsymbol{X}_1}, \boldsymbol{\beta}_{\boldsymbol{X}_2})^2},
\end{aligned} \tag{12}$$

which is equal to θ_1 if the InSIDE assumption holds with respect to $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ and $\boldsymbol{\beta}_{\boldsymbol{X}_2}$. As more risk factors with correlated sets of association parameters with $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ are included in the multivariable MR-Egger model, additional terms will be added to the bias term in Equation 12, and the InSIDE assumption must hold for these additional risk factors to obtain a consistent estimate of θ_1.

The variance of the multivariable MR-Egger estimate $\hat{\theta}_{1ME}$ will be heavily influenced by the denominator in the bias term of Equation 12. As $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ and $\boldsymbol{\beta}_{\boldsymbol{X}_2}$ become more highly correlated, the standard error of the causal estimate $\hat{\theta}_{1ME}$ will increase, and in some circumstances, the estimate from multivariable MR-Egger will be less precise than the estimate from univariable MR-Egger. The precision of the causal estimates from multivariable MR-Egger and univariable MR-Egger is discussed further in the Web Appendix.

2.4 | Advantages of multivariable MR-Egger and comparison with univariable MR-Egger

The bias for the causal estimate from univariable MR-Egger $\hat{\theta}_{UE}$ depends on the weighted covariance between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_{\boldsymbol{X}_1}$, where

$$\alpha_j = \alpha'_j + \sum_{i=2}^{K} \theta_i \beta_{X_{ij}}. \tag{13}$$

The expression in Equation 13 follows from the multivariable framework outlined in Equation 5, where the direct effect for univariable MR-Egger has been decomposed into the residual direct effect α'_j of multivariable MR-Egger and the indirect effects via each risk factor. The residual direct effect α'_j will be altered with each additional risk factor included in the multivariable MR-Egger model. If these additional risk factors are causally associated with the outcome ($\theta_k \neq 0$), then α'_j will consist of fewer components. It seems likely that the InSIDE assumption would be easier to satisfy for multivariable

MR-Egger than its univariable counterpart as the direct effect for univariable MR-Egger consists of unmeasured and measured pleiotropy.

If the $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ parameters are independent of the $\boldsymbol{\beta}_{\boldsymbol{X}_k}$ parameters for all $k = 2, 3, \ldots, K$, then the second term in Equation 13 (the measured direct effect) does not contribute to the value of $\text{cov}_w(\boldsymbol{\alpha}, \boldsymbol{\beta}_{\boldsymbol{X}_1})$. Under this scenario, bias for the univariable and multivariable MR-Egger estimates depends on the same covariance term $\text{cov}_w(\boldsymbol{\alpha}', \boldsymbol{\beta}_{\boldsymbol{X}_1})$. As a consequence, the estimates of the causal effects from univariable MR-Egger $\hat{\theta}_{UE}$ and multivariable MR-Egger $\hat{\theta}_{1ME}$ will be asymptotically the same. In this case, multivariable MR-Egger may improve precision of the causal estimate but will not affect the asymptotic bias.

When the $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ parameters are correlated with at least one of the sets of $\boldsymbol{\beta}_{\boldsymbol{X}_k}$ parameters for $k = 2, 3, \ldots, K$, the second term in Equation 13 now contributes to the value of $\text{cov}_w(\boldsymbol{\alpha}, \boldsymbol{\beta}_{\boldsymbol{X}_1})$. The InSIDE assumption for univariable MR-Egger will therefore be automatically violated as the weighted covariance between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ will not equal zero, resulting in biased causal estimates of θ_1. If the InSIDE assumption holds for multivariable MR-Egger, and $\boldsymbol{\beta}_{\boldsymbol{X}_k}$ are included in the analysis model, then $\hat{\theta}_{1ME}$ will still be a consistent estimate of θ_1. Hence, in this case, multivariable MR-Egger should result in reduced bias compared with univariable MR-Egger.

2.5 | Orientation of the genetic variants

Genetic associations represent the average change in the risk factor or the outcome per additional copy of the reference allele. There is no biological rationale why associations should be expressed with respect to either the major (wildtype) or the minor (variant) allele. In the univariable and multivariable IVW methods, the estimate is not affected by the choice of orientation, as the intercept is fixed at zero. However, in the univariable and multivariable MR-Egger methods, changing the orientation of the variant affects the intercept term and the causal estimate as the orientation affects the definition of the pleiotropy terms α_j and α'_j. Consequently, for each choice of orientation, there is a different version of the InSIDE assumption.

To ensure that the MR-Egger analysis does not depend on the reported reference alleles, Bowden et al suggested the genetic variants in univariable MR-Egger be orientated so the direction of association with the risk factor is either positive for all variants or negative for all variants.[9] However, this may not be possible for multivariable MR-Egger as the same reference allele must be used for associations with each risk factor and with the outcome. We suggest that the variants should be orientated with respect to their associations with the risk factor of primary interest, although we would recommend a sensitivity analysis considering different orientations if multiple risk factors are of interest. If the genetic variants are all valid instruments, then directional pleiotropy should not be detected with respect to any orientation.

3 | EXAMPLE: CAUSAL EFFECT OF HDL-C ON CHD RISK

The effects of high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglycerides on the risk of coronary heart disease (CHD) have been investigated by numerous MR studies.[21] For HDL-C, univariable MR suggested a causally protective role against CHD risk, whereas univariable MR-Egger provided no evidence of a causal effect and the test for directional pleiotropy was statistically significant at the 5% level.[8] A null causal effect for HDL-C was also reported from a multivariable MR analysis that included LDL-C and triglycerides using the multivariable IVW method,[7] although a small but protective causal effect was estimated in a further multivariable MR analysis using a wider range of 185 genetic variants.[22]

We investigate the causal effect of HDL-C on CHD risk further using the multivariable MR-Egger method. We consider the 185 genetic variants having known association with at least one of HDL-C, LDL-C, and triglycerides at GWAS significance in 188 578 participants reported by the Global Lipids Genetics Consortium.[23] The point estimates for the associations between these genetic variants and lipids were taken from Do et al.[24] The CARDIoGRAMplusC4D consortium consisting of 60 801 cases and 123 504 controls was used to obtain the estimates of the association between the variants and CHD risk.[25] The IVW and MR-Egger methods were applied to the data under univariable and multivariable frameworks as described in Section 2. For the univariable IVW and MR-Egger methods, the models were fitted using 2 sets of variants: firstly using all 185 variants; and secondly using all variants associated with HDL-C at GWAS level of significance. The genetic variants were orientated with respect to the risk increasing allele for HDL-C. These analyses differ from those provided in Burgess et al and Do et al as they use summarized data from different versions of the CARDIoGRAMplusC4D study[22,24]; here, we use associations from the 2015 data release.[25]

TABLE 1 Log causal odds ratios (95% confidence intervals) for coronary heart disease per standard deviation increase in HDL-C, with 2-sided *P*-values. Estimates of the intercept are given in univariable and multivariable MR-Egger

	Causal Estimate			MR-Egger Intercept Test		
	$\hat{\theta}_{\text{HDL-C}}$ (CI)	se($\hat{\theta}_{\text{HDL-C}}$)	*P*-value	$\hat{\theta}_{0E}$	se($\hat{\theta}_{0E}$)	*P*-value
Univariable IVW						
All variants	−0.130 (−0.227, −0.033)	0.049	0.009	-	-	-
Reduced set of variants[a]	−0.114 (−0.211, −0.017)	0.049	0.022	-	-	-
Univariable MR-Egger						
All variants	−0.016 (−0.138, 0.106)	0.062	0.800	−0.007	0.002	0.004
Reduced set of variants[a]	0.067 (−0.070, 0.204)	0.069	0.332	−0.012	0.004	0.001
Multivariable IVW	−0.039 (−0.123, 0.045)	0.042	0.359	-	-	-
Multivariable MR-Egger	0.036 (−0.063, 0.134)	0.050	0.477	−0.005	0.002	0.008

Abbreviations: CI, confidence interval; HDL-C, high-density lipoprotein cholesterol; IVW, inverse-variance weighted; MR, Mendelian randomization; SE, standard error.
[a]95 variants associated with HDL-C at a genome-wide level of significance (*P*-value$< 5 \times 10^{-8}$).

TABLE 2 Causal log odds ratios (95% confidence intervals) for coronary heart disease per standard deviation increase in HDL-C, LDL-C, and triglycerides from multivariable IVW and multivariable MR-Egger. Estimates from multivariable MR-Egger are presented from 3 models where the reference allele is the risk increasing allele for HDL-C, LDL-C, or triglycerides. Estimates of the intercept are given for multivariable MR-Egger

	Causal Estimates			MR-Egger Intercept
	$\tilde{\theta}_{\text{HDL-C}}$	$\hat{\theta}_{\text{LDL-C}}$	$\hat{\theta}_{\text{TG}}$	$\tilde{\theta}_{0E}$
Multivariable IVW	−0.039 (−0.123, 0.045)	0.375 (0.292, 0.457)	0.173 (0.063, 0.283)	-
Multivariable MR-Egger				
Orientation with respect to[a]:				
HDL-C	**0.036 (−0.063, 0.134)**	0.378 (0.297, 0.458)	0.136 (0.024, 0.247)	−0.005 (−0.009, −0.001)
LDL-C	−0.034 (−0.118, 0.049)	**0.420 (0.318, 0.522)**	0.194 (0.081, 0.308)	−0.003 (−0.007, 0.001)
TG	−0.018 (−0.102, 0.066)	0.350 (0.267, 0.433)	**0.083 (−0.045, 0.211)**	0.005 (0.001, 0.009)

Abbreviations: HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; MR, Mendelian randomization; TG, triglycerides.
[a]Alleles orientated for all genetic associations with respect to the risk increasing allele for HDL-C, LDL-C, or triglycerides.

The univariable IVW method suggested a significant protective effect of HDL-C for both sets of variants with a causal odds ratio of 0.88 (95% CI: 0.80-0.97) for all variants (Table 1). This estimate attenuated to the null in the univariable MR-Egger method (0.98, 95% CI: 0.87-1.11) with evidence of directional pleiotropy (*P*-value = 0.004). The causal odds ratios from multivariable IVW (0.96, 95% CI: 0.89-1.05) and multivariable MR-Egger (1.04, 95% CI: 0.94-1.14) had opposite directions of association, with both analyses indicating that HDL-C is not causally associated with CHD risk. The significant result for directional pleiotropy in the multivariable MR-Egger method suggests that LDL-C and triglycerides do not fully explain the direct effects of the genetic variants on the outcome, suggesting that there is still residual pleiotropy via other unmeasured risk factors.

3.1 | Varying the orientation of the genetic variants

As a sensitivity analysis, the multivariable MR-Egger method was reperformed with the genetic variants orientated with respect to the risk increasing alleles for LDL-C and triglycerides.

The causal estimates for HDL-C, LDL-C, and triglycerides from multivariable MR-Egger when the variants were orientated with respect to HDL-C, LDL-C or triglycerides are presented in Table 2. Estimates of the MR-Egger intercept are also provided for the three models. To allow for comparisons between the multivariable methods, the causal estimates from multivariable IVW are included in Table 2. The causal estimates in bold follow the recommendation outlined in Section 2.5 that the genetic variants should be orientated with respect to the risk factor-increasing allele for the risk factor of interest.

All of the causal odds ratios for HDL-C from the multivariable MR-Egger models indicated that HDL-C is not causally associated with CHD risk. Significant adverse effects of LDL-C on CHD risk were reported from the multivariable IVW

(1.45, 95% CI: 1.34-1.58) and multivariable MR-Egger (1.52, 95% CI: 1.37-1.69) methods. Orientating the variants with respect to the risk increasing alleles for HDL-C and triglycerides had little impact on the causal estimates for LDL-C from multivariable MR-Egger. The multivariable IVW method suggested a significant adverse effect of triglycerides on CHD risk with a causal odds ratio of 1.19 (95% CI: 1.07, 1.33), this estimate was attenuated to the null in the multivariable MR-Egger method (1.09, 95% CI: 0.96, 1.23). The causal odds ratios for triglycerides remained significant, however, when the variants were orientated with respect to HDL-C and LDL-C in the multivariable MR-Egger models.

Since the orientation of the genetic variants affects the interpretation of the direct effect, and the definition of the InSIDE assumption, the MR-Egger intercept will vary between different orientations. In this example, the MR-Egger intercept differed from zero when the variants were orientated with respect to HDL-C and triglycerides, yet there was no evidence of directional pleiotropy or the InSIDE assumption being violated when the variants were orientated with respect to LDL-C.

4 | SIMULATION STUDY

To assess the merits of using multivariable MR-Egger over multivariable IVW and univariable MR-Egger in realistic settings, we perform a simulation study. Univariable and multivariable MR-Egger will be compared with respect to the consistency of the causal estimates and statistical power to detect the causal effect. The setup of the simulation study corresponds to the applied example in Section 3 and will be considered under 2 broad scenarios: (1) $\boldsymbol{\beta}_{\boldsymbol{X_k}}$ are generated independently for all $k = 1, 2, \ldots, K$; and (2) $\boldsymbol{\beta}_{\boldsymbol{X_k}}$ are correlated for all $k = 1, 2, \ldots, K$.

We simulated summarized level data for 185 genetic variants indexed by $j = 1, 2, \ldots, J$ for 3 risk factors (X_1, X_2, X_3) and an outcome Y from the following data-generating model:

$$\begin{aligned}
\begin{pmatrix} \beta_{X_{1j}} \\ \beta_{X_{2j}} \\ \beta_{X_{3j}} \end{pmatrix} &\sim \mathcal{N}_3\left(\begin{pmatrix} 0.08 \\ 0.03 \\ -0.05 \end{pmatrix}, \begin{pmatrix} {\sigma_1}^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & {\sigma_2}^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & {\sigma_3}^2 \end{pmatrix}\right) \\
\beta_{Yj} &= \alpha'_j + \theta_1|\beta_{X_{1j}}| + \theta_2\beta_{X_{2j}} + \theta_3\beta_{X_{3j}} + \epsilon_j \\
\epsilon_j &\sim \mathcal{N}(0, 1) \\
\alpha'_j &\sim \mathcal{N}(\mu, 0.004).
\end{aligned} \tag{14}$$

The primary objective was to estimate θ_1, with the causal effects set to: $\theta_1 = 0$ (null causal effect) or $\theta_1 = 0.3$ (positive causal effect); $\theta_2 = 0.1$; and $\theta_3 = -0.3$. The data were simulated to consider the following four scenarios:

1. No pleiotropy ($\alpha'_j = 0$ for all j), InSIDE assumption automatically satisfied;
2. Balanced pleiotropy ($\mu = 0$), InSIDE assumption satisfied;
3. Directional pleiotropy ($\mu = 0.01, 0.05$ or 0.1), InSIDE assumption satisfied;
4. Directional pleiotropy ($\mu = 0.01, 0.05$ or 0.1), InSIDE assumption violated.

When the InSIDE assumption for multivariable MR-Egger was satisfied, α'_j and $\beta_{X_{1j}}$ were drawn from independent distributions, and when it was violated, they were drawn from a multivariate normal distribution with $\text{cor}(\boldsymbol{\alpha'}, \boldsymbol{\beta}_{\boldsymbol{X_1}}) = 0.3$. The above 4 scenarios were applied to the simulated data when $\boldsymbol{\beta}_{\boldsymbol{X_k}}$ were generated independently for all k, with the parameters in the covariance matrix set to: $\sigma_1^2 = 0.03$; $\sigma_2^2 = 0.02$; $\sigma_3^2 = 0.04$; and $\rho_{12} = \rho_{13} = \rho_{23} = 0$. The 4 scenarios were repeated when $\boldsymbol{\beta}_{\boldsymbol{X_k}}$ were correlated for all k ($\rho_{12} = 0.2$, $\rho_{13} = -0.3$, $\rho_{23} = 0.1$). The mean F-statistics were greater than 200 and I^2 statistics greater than 99% in each scenario; values are provided in Web Tables A1 and A2. In total, data were simulated for 32 different choices of parameters.

To ensure the direction of association between G_j and X_1 was the same for all j variants, the absolute value of the genetic associations with X_1 ($|\beta_{X_{1j}}|$) were used to generate β_{Y_j} (Equation (14)). It was assumed that $\beta_{X_{kj}}$ (for all k) and β_{Y_j} had the same reference allele and the genetic variants were uncorrelated. The multivariable IVW, univariable MR-Egger, and multivariable MR-Egger methods were applied to the simulated datasets. The weights for the multivariable IVW and multivariable MR-Egger are given by Equation 15, while Equation 16 contains the weights for univariable MR-Egger:

$$\text{se}(\beta_{Y_j})^{-2} = ({\epsilon_j}^2 + {\sigma_{\alpha'}}^2)^{-1}, \tag{15}$$

$$\text{se}(\beta_{Y_j})^{-2} = ({\epsilon_j}^2 + {\sigma_{\alpha'}}^2 + {\theta_2}^2{\sigma_2}^2 + {\theta_3}^2{\sigma_3}^2)^{-1}. \tag{16}$$

4.1 | Results

The results from the simulation study using 10 000 simulated datasets are presented in Table 3 ($\boldsymbol{\beta}_{\boldsymbol{X}_k}$ generated independently) and Table 4 ($\boldsymbol{\beta}_{\boldsymbol{X}_k}$ correlated). For each scenario, the mean estimate, the mean standard error, and the statistical power to detect a null or positive causal effect at a nominal 5% significance level are presented in Tables 3 and 4 for the multivariable IVW, univariable MR-Egger, and multivariable MR-Egger methods. For univariable and multivariable MR-Egger, the statistical power of the MR-Egger intercept test is also provided.

$\boldsymbol{\beta}_{\boldsymbol{X}_k}$ generated independently: In scenarios 1 and 2 (no and balanced pleiotropy), estimates from all methods were unbiased, and those from the multivariable IVW method were the most precise. In scenarios 3 and 4 (directional pleiotropy), estimates from the multivariable IVW method were biased, with the magnitude of bias increasing as the average value of $\boldsymbol{\alpha}'$ increased from 0.01 to 0.1. In scenario 3 (InSIDE satisfied), estimates from the univariable and multivariable MR-Egger methods were unbiased, whereas in scenario 4 (InSIDE violated), they were biased. Although the causal estimates for both multivariable IVW and multivariable MR-Egger were biased under scenario 4, the magnitude of bias was less for multivariable MR-Egger, with the exception of when α_j' was generated from $\mathcal{N}(0.01, 0.004)$. Precision and power to detect a causal effect were always better for the multivariable MR-Egger method than univariable

TABLE 3 Performance of multivariable IVW, univariable MR-Egger, and multivariable MR-Egger with respect to $\hat{\theta}_1$ for a null ($\theta_1 = 0$) and positive ($\theta_1 = 0.3$) causal effect where $\boldsymbol{\beta}_{\boldsymbol{X}_k}$ are generated independently for all k. All tests were performed at the 5% level of significance

	Multivariable IVW		Univariable MR-Egger			Multivariable MR-Egger		
	Mean $\hat{\theta}_1$ (mean SE)	Power, %	Mean $\hat{\theta}_1$ (mean SE)	Power, % Intercept	Causal	Mean $\hat{\theta}_1$ (mean SE)	Power, % Intercept	Causal
Null causal effect: $\theta_1 = 0$								
1. No pleiotropy, InSIDE satisfied								
	0.000 (0.045)	3.8	−0.002 (0.158)	9.1	4.7	0.000 (0.084)	3.7	4.1
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha_j' \sim \mathcal{N}(0,0.004)$	-0.001 (0.100)	4.7	−0.001 (0.187)	7.8	4.7	0.000 (0.165)	4.6	4.6
3. Directional pleiotropy, InSIDE satisfied								
$\alpha_j' \sim \mathcal{N}(0.01,0.004)$	0.041 (0.100)	6.7	−0.003 (0.187)	12.2	4.3	−0.002 (0.165)	5.9	4.5
$\alpha_j' \sim \mathcal{N}(0.05,0.004)$	0.210 (0.100)	55.3	0.002 (0.187)	49.2	4.6	0.002 (0.166)	36.3	4.6
$\alpha_j' \sim \mathcal{N}(0.1,0.004)$	0.417 (0.102)	97.4	0.000 (0.187)	91.6	4.3	0.001 (0.165)	88.0	4.6
4. Directional pleiotropy, InSIDE violated								
$\alpha_j' \sim \mathcal{N}(0.01,0.004)$	0.074 (0.100)	12.3	0.089 (0.187)	6.7	7.6	0.088 (0.165)	4.3	8.4
$\alpha_j' \sim \mathcal{N}(0.05,0.004)$	0.240 (0.100)	67.2	0.089 (0.187)	34.1	7.8	0.088 (0.165)	21.1	8.8
$\alpha_j' \sim \mathcal{N}(0.1,0.004)$	0.450 (0.101)	98.6	0.088 (0.187)	84.1	7.6	0.088 (0.165)	78.7	8.7
Positive causal effect: $\theta_1 = 0.3$								
1. No pleiotropy, InSIDE satisfied								
	0.300 (0.044)	98.9	0.300 (0.157)	9.3	50.1	0.300 (0.084)	4.3	87.3
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha_j' \sim \mathcal{N}(0,0.004)$	0.301 (0.100)	84.6	0.303 (0.187)	7.5	38.2	0.302 (0.166)	4.9	46.4
3. Directional pleiotropy, InSIDE satisfied								
$\alpha_j' \sim \mathcal{N}(0.01,0.004)$	0.343 (0.100)	91.5	0.300 (0.187)	12.8	36.8	0.299 (0.165)	6.0	45.8
$\alpha_j' \sim \mathcal{N}(0.05,0.004)$	0.509 (0.100)	99.7	0.300 (0.188)	50.6	37.3	0.299 (0.166)	37.1	46.1
$\alpha_j' \sim \mathcal{N}(0.1,0.004)$	0.716 (0.102)	100.0	0.300 (0.187)	91.1	37.1	0.299 (0.166)	87.9	46.1
4. Directional pleiotropy, InSIDE violated								
$\alpha_j' \sim \mathcal{N}(0.01,0.004)$	0.374 (0.099)	94.3	0.390 (0.187)	6.6	56.4	0.389 (0.165)	4.6	65.8
$\alpha_j' \sim \mathcal{N}(0.05,0.004)$	0.539 (0.100)	99.8	0.388 (0.187)	34.4	55.6	0.387 (0.165)	21.5	65.5
$\alpha_j' \sim \mathcal{N}(0.1,0.004)$	0.747 (0.101)	100.0	0.383 (0.187)	84.7	55.1	0.384 (0.165)	78.3	65.2

Abbreviations: InSIDE, Instrument Strength Independent of Direct Effect; IVW, inverse-variance weighted; MR, Mendelian randomization; SE, standard error.

TABLE 4 Performance of multivariable IVW, univariable MR-Egger, and multivariable MR-Egger with $\boldsymbol{\beta}_{\boldsymbol{X}_k}$ being correlated for all k

	Multivariable IVW		Univariable MR-Egger			Multivariable MR-Egger		
	Mean $\hat{\theta}_1$ (mean SE)	Power, %	Mean $\hat{\theta}_1$ (mean SE)	Power, % Intercept	Causal	Mean $\hat{\theta}_1$ (mean SE)	Power, % Intercept	Causal
Null causal effect: $\theta_1 = 0$								
1. No pleiotropy, InSIDE satisfied								
	0.000 (0.047)	4.0	0.099 (0.157)	4.3	10.1	0.000 (0.086)	4.4	4.6
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha_j' \sim \mathcal{N}(0,0.004)$	−0.001 (0.104)	4.7	0.093 (0.187)	4.5	7.4	−0.003 (0.169)	4.6	4.4
3. Directional pleiotropy, InSIDE satisfied								
$\alpha_j' \sim \mathcal{N}(0.01,0.004)$	0.043 (0.104)	7.0	0.099 (0.187)	5.8	8.0	0.001 (0.169)	5.9	4.8
$\alpha_j' \sim \mathcal{N}(0.05,0.004)$	0.213 (0.105)	52.7	0.095 (0.187)	33.3	7.6	0.000 (0.169)	37.2	4.5
$\alpha_j' \sim \mathcal{N}(0.1,0.004)$	0.426 (0.107)	96.3	0.096 (0.187)	84.5	7.6	−0.001 (0.169)	89.2	4.6
4. Directional pleiotropy, InSIDE violated								
$\alpha_j' \sim \mathcal{N}(0.01,0.004)$	0.062 (0.104)	9.5	0.184 (0.187)	4.6	17.9	0.078 (0.169)	4.7	7.6
$\alpha_j' \sim \mathcal{N}(0.05,0.004)$	0.235 (0.104)	62.1	0.187 (0.187)	20.5	18.3	0.082 (0.169)	22.3	7.5
$\alpha_j' \sim \mathcal{N}(0.1,0.004)$	0.448 (0.106)	97.9	0.181 (0.187)	73.3	17.8	0.077 (0.169)	80.3	7.2
Positive causal effect: $\theta_1 = 0.3$								
1. No pleiotropy, InSIDE satisfied								
	0.300 (0.047)	98.7	0.395 (0.158)	4.4	70.8	0.299 (0.087)	3.9	86.2
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha_j' \sim \mathcal{N}(0,0.004)$	0.300 (0.104)	81.5	0.399 (0.187)	4.4	58.0	0.301 (0.169)	4.6	44.4
3. Directional pleiotropy, InSIDE satisfied								
$\alpha_j' \sim \mathcal{N}(0.01,0.004)$	0.342 (0.104)	89.4	0.395 (0.187)	6.4	57.4	0.301 (0.169)	5.9	44.4
$\alpha_j' \sim \mathcal{N}(0.05,0.004)$	0.513 (0.105)	99.4	0.394 (0.187)	33.0	57.4	0.296 (0.169)	38.0	43.4
$\alpha_j' \sim \mathcal{N}(0.1,0.004)$	0.729 (0.107)	100.0	0.400 (0.187)	83.5	58.2	0.304 (0.169)	88.6	45.5
4. Directional pleiotropy, InSIDE violated								
$\alpha_j' \sim \mathcal{N}(0.01,0.004)$	0.365 (0.104)	92.1	0.489 (0.187)	4.2	74.0	0.382 (0.169)	4.6	63.2
$\alpha_j' \sim \mathcal{N}(0.05,0.004)$	0.535 (0.104)	99.7	0.486 (0.187)	20.3	72.9	0.382 (0.169)	21.1	63.2
$\alpha_j' \sim \mathcal{N}(0.1,0.004)$	0.749 (0.106)	100.0	0.488 (0.187)	72.5	73.4	0.381 (0.169)	79.6	62.8

Abbreviations: InSIDE, Instrument Strength Independent of Direct Effect; IVW, inverse-variance weighted; MR, Mendelian randomization; SE, standard error.

MR-Egger, although the univariable MR-Egger method detected directional pleiotropy more often. The average value of $\boldsymbol{\alpha'}$ had no impact on the degree of bias for univariable or multivariable MR-Egger.

$\boldsymbol{\beta}_{\boldsymbol{X}_k}$ correlated: Bias for the multivariable IVW method was present in scenarios 3 and 4 only, as in the independently generated setting. In this setting, the InSIDE assumption for univariable MR-Egger was violated for all 4 scenarios, resulting in biased point estimates of θ_1. However, the multivariable InSIDE assumption was satisfied for scenarios 1, 2, and 3, and so causal estimates from multivariable MR-Egger were unbiased. When the multivariable InSIDE assumption was violated (scenario 4) the estimates from multivariable MR-Egger were biased, yet the magnitude of bias was less compared with univariable MR-Egger as $|\text{cov}(\boldsymbol{\alpha'}, \boldsymbol{\beta}_{\boldsymbol{X}_1})| < |\text{cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}_{\boldsymbol{X}_1})|$.

4.2 | Causal relationships between the risk factors

The simulations performed in Section 4.1 assumed that the effect of each risk factor on the outcome is not mediated through another risk factor. There may be circumstances where causal relationships between risk factors are biologically plausible. Burgess et al illustrated that the multivariable IVW method estimates the direct causal effects (θ_k) of each risk factor on the outcome, irrespective of whether causal relationships between the risk factors exist.[7]

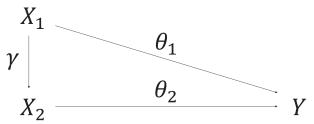


FIGURE 3 Causal directed acyclic graph illustrating the causal relationships between the 2 risk factors X_1 and X_2, and outcome Y. The causal effect of X_1 on X_2 is γ, and the direct causal effect of the risk factor X_k on the outcome Y is θ_k. The total causal effect of X_1 on Y is $\theta_1 + \gamma\theta_2$, consisting of the direct effect (θ_1) and the indirect effect via X_2 ($\gamma\theta_2$). U_k represents the set of variables that confound the associations between X_k and Y

In the applied example of the paper, there may also be deterministic dependencies between the risk factors. LDL-C is rarely measured directly but is estimated from measurements of total cholesterol, triglycerides, and HDL-C via the Friedewald equation as total cholesterol minus HDL-C minus 0.2 times triglycerides (assuming all measurements in mg/dL).[26] It has previously been shown that the coefficient for LDL-C is the same as the coefficient for non-HDL-C (calculated as total cholesterol minus HDL-C) in a regression model including HDL-C and triglycerides (see Appendix 2 in the paper by Di Angelantonio et al).[27] However, the coefficient for triglycerides will change, as the non-HDL-C measure contains more triglycerides than the LDL-C measure. Hence, in the case that there are deterministic relationships between the risk factors, effect estimates may change as the choice of risk factors varies due to their interpretation as direct effects conditional on other risk factors in the regression model.

We performed additional simulations to investigate the behaviour of the multivariable MR-Egger method when X_2 is causally dependent on X_1, and the causal effect of X_1 on X_2 is γ (Figure 3). The total causal effect of X_1 on Y is $\theta_1 + \gamma\theta_2$, consisting of the direct effect (θ_1) and the indirect effect via X_2 ($\gamma\theta_2$). See the Web Appendix for more details on the data generating model.

4.2.1 | Results

The results from the additional simulations are provided in Web Table A3 and Web Table A4. In scenarios where there was no bias in the original set of simulations, the multivariable IVW and multivariable MR-Egger methods consistently estimated the direct effect of X_1 on Y (θ_1), while the univariable MR-Egger method consistently estimated the total causal effect of X_1 on Y ($\theta_1 + \gamma\theta_2$). Compared to the results in Section 4.1, precision and power to detect a causal effect were reduced for the multivariable IVW and multivariable MR-Egger methods. This reduction in power was anticipated since the multivariable models condition on the mediator along a causal pathway, which is known to decrease power to detect a causal effect.[28]

5 | DISCUSSION

In this paper, we have extended univariable MR-Egger to the multivariable setting and outlined the assumptions required to obtain consistent causal estimates in the presence of directional pleiotropy. Multivariable MR-Egger should be viewed as a sensitivity analysis to provide robustness against both measured and unmeasured pleiotropy and to strengthen the evidence from the original MR analysis. If the causal estimate from multivariable MR-Egger is substantially different from the estimate obtained in the original analysis, then further investigation into the causal finding and the potential for pleiotropy is required.

The simulation study has highlighted the benefits of using multivariable MR-Egger over its univariable counterpart. This is particularly true when the associations of the genetic variants with the risk factor of interest are associated with genetic associations with at least one of the risk factors (measured pleiotropy). Under this scenario, the InSIDE assumption for univariable MR-Egger is likely to be violated, leading to biased causal estimates. Multivariable MR-Egger will, however, produce consistent causal estimates if the InSIDE assumption for multivariable MR-Egger is satisfied. Although the estimates from univariable and multivariable MR-Egger are asymptotically the same when genetic associations with each risk factor are all independent, multivariable MR-Egger should also have greater power to detect a causal effect when the InSIDE assumption is satisfied. Given these advantages, and the sensitivity of the multivariable IVW method to directional pleiotropy, we believe that multivariable MR-Egger should be considered as an important sensitivity analysis for a MR study.

5.1 | Multivariable by design, or multivariable as a sensitivity analysis?

There are 2 possible scenarios where multivariable MR-Egger may be used as a sensitivity analysis: either the primary analysis is considered to be multivariable by design, or a multivariable framework is only considered as part of the sensitivity analysis. The first case should be motivated by biological evidence where the set of risk factors are known to be associated with common genetic variants, such as lipid fractions. Under this scenario, multivariable IVW should be used as the primary analysis method with multivariable MR-Egger providing robustness against directional pleiotropy as a sensitivity analysis.

In the second scenario, where there is a lack of biological evidence to suggest a multivariable framework, univariable IVW would generally be considered as the primary analysis method and univariable MR-Egger as the main sensitivity analysis. However, if the genetic variants are associated with other risk factors, multivariable MR-Egger could also be used as a sensitivity analysis as its assumptions are more likely to be satisfied and it may have greater power to detect a causal effect than univariable MR-Egger. An example of the use of multivariable MR as a sensitivity analysis is an MR study on plasma urate concentrations and CHD risk.[29] To account for measured and unmeasured pleiotropic associations of the genetic variants, the authors performed the multivariable IVW and univariable MR-Egger methods as sensitivity analyses. This investigation may have benefited from performing the multivariable MR-Egger method to simultaneously account for both measured and unmeasured pleiotropic associations.

5.2 | InSIDE assumption and orientation of genetic variants

The validity of multivariable MR-Egger and its ability to estimate consistent causal effects is dependent upon the InSIDE assumption being satisfied. While it is not possible to determine whether the InSIDE assumption has been violated, we believe it is more likely to hold for multivariable MR-Egger then univariable MR-Egger. When the $\boldsymbol{\beta}_{\boldsymbol{X}_1}$ parameters are associated with at least one of the sets of $\boldsymbol{\beta}_{\boldsymbol{X}_k}$ parameters for $k = 2, 3, \ldots, K$, the InSIDE assumption for univariable MR-Egger is automatically violated and causal estimates from the method will be inconsistent. The direct effects of the genetic variants on the outcome will consist of fewer components for multivariable MR-Egger compared to its univariable counterpart, making it more plausible that the InSIDE assumption will hold for multivariable MR-Egger.

The recommendation of orientating the genetic variants in multivariable MR-Egger to the risk factor-increasing or risk factor-decreasing allele for the risk factor of interest may be considered arbitrary. While we accept this limitation, we would argue that it brings consistency to the results. This recommendation may result in the analysis being performed up to K times to obtain the causal estimates for all K risk factors. The orientation of the genetic variants will also affect the interpretation of the direct effect, thereby altering the InSIDE assumption. This may result in the MR-Egger intercept estimate varying between different orientations. This was seen in the applied example where the intercept term was non-significant when the alleles were orientated with respect to LDL-C, and significant when orientated with respect to HDL-C and trigclyercides.

5.3 | Linearity and homogeneity assumptions

Throughout this paper, we have assumed linearity and homogeneity (no effect modification) of the causal effects of the risk factors on the outcome, and of the associations between the genetic variants with the risk factors and with the outcome. If the assumptions of linearity and homogeneity are violated then the methods discussed in this paper still provide a valid test for the null hypothesis of whether the risk factor is causally associated with the outcome.[12] The causal estimate, however, would not have a literal interpretation if the assumptions were violated.[30] Although linearity and homogeneity are strong assumptions, the effect of genetic variants on the risk factor and outcome tend to be limited to a small range, which may make the assumptions of linearity and homogeneity more reasonable in an MR analysis.

The multivariable models have assumed that the risk factors do not have causal effects on each other. The additional simulation study has illustrated that the multivariable MR-Egger method estimates the direct causal effects of the risk factors on the outcome, irrespective of whether the risk factors are causally related. There was, however, a reduction in precision and power to detect the causal effect for multivariable MR-Egger when a causal relationship between the risk factors was present. Conversely, univariable MR-Egger will produce consistent causal estimates of the total effect if the InSIDE assumption for univariable MR-Egger is satisfied.

5.4 | Implication for future research

The paper by Helgadottir et al highlights the importance and need to develop sensitivity analyses for multivariable MR.[10] This is particularly relevant given the recent advances in high-throughput phenotyping which has led to the introduction of "-omics" data such as metabolomics, genomics, and proteomics.[31] Genome-wide analyses of high-dimensional "-omics" data are becoming more popular,[32,33] yet few MR analyses have been performed using these datasets.[21] As summarized data from large consortia become more accessible, the opportunities to use MR on high-dimensional datasets will only increase. Methods such as multivariable MR-Egger will be valuable to investigate the causal effects of multiple related phenotypes with shared genetic predictors.

Bowden et al have shown that uncertainty in the associations between the genetic variants and the risk factor in univariable MR-Egger can lead to attenuation towards the null when a causal effect exists between the risk factor and the outcome.[34] This attenuation is approximately equal to the I^2 statistic from meta-analysis of the weighted associations with the exposure $\hat{\beta}_{X_j}\text{se}(\hat{\beta}_{Y_j})^{-1}$, with standard errors $\text{se}(\hat{\beta}_{X_j})\text{se}(\hat{\beta}_{Y_j})^{-1}$.[34] Since the mean I^2 statistics for the simulation study in this paper were close to 100%, there was no substantial bias in the causal estimates due to uncertainty in the genetic associations for either the univariable or multivariable MR-Egger methods. However, it is unclear whether uncertainty in the genetic associations with the risk factors would always lead to the attenuation of the causal estimates for the multivariable MR-Egger method. Further research is required to investigate this.

Throughout the paper, we have assumed that the genetic variants are uncorrelated (not in linkage disequilibrium). This assumption, and the requirement for further methodological development, is discussed in the Web Appendix.

ACKNOWLEDGEMENTS

Jessica Rees is supported by the British Heart Foundation (grant number FS/14/59/31282). Stephen Burgess is supported by Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 204623/Z/16/Z).

ORCID

Jessica M. B. Rees http://orcid.org/0000-0002-1495-2470
Stephen Burgess http://orcid.org/0000-0001-5365-8760

REFERENCES

1. Smith GD, Ebrahim S. Mendelian randomization: Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1-22.
2. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008;27(8):1133-1163.
3. Sleiman PM, Grant SF. Mendelian randomization in the era of genomewide association studies. *Clin Chem*. 2010;56(5):723-728.
4. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol*. 2013;42(5):1497-1501.
5. Freeman G, Cowling BJ, Schooling CM. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *Int J Epidemiol*. 2013;42(4):1157-1163.
6. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol*. 2013;42(4):1134-1144.
7. Burgess S, Thompson SG. Multivariable Mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol*. 2015;181(4):251-260.
8. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304-314.
9. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512-525.
10. Helgadottir A, Gretarsdottir S, Thorleifsson G, et al. Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. *Nat Genet*. 2016;48(6):634-639.
11. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG. Using published data in Mendelian randomization: A blueprint for efficient identification of causal risk factors. *Eur J Epidemiol*. 2015;30(7):543-552.
12. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods. *Stat Med*. 2016;35(11):1880-1906.
13. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol*. 2000;29(4):722-729.

14. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res*. 2007;16(4):309-330.
15. Johnson T. Efficient calculation for multi-SNP genetic risk scores, The Comprehensive R Archive Network; 2013. http://cran.r-project.org/web/packages/gtx/vignettes/ashg2012.pdf. Accessed June 14, 2016.
16. Angrist JD, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *J Am Stat Assoc*. 1995;90(430):431-442.
17. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658-665.
18. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: A comparison of methods. *Stat Med*. 1999;18(20):2693-2708.
19. Kolesár M, Chetty R, Friedman J, Glaeser E, Imbens G. Identification and inference with many invalid instruments. *J Bus Econ Stat*. 2015;33(4):474-484. http://doi.org.10.1080/07350015.2014.978175
20. Burgess S, Dudbridge F, Thompson SG. Re: "Multivariable Mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects". *Am J Epidemiol*. 2015;181(4):290-291.
21. Burgess S, Harshfield E. Mendelian randomization to assess causal effects of blood lipids on coronary heart disease: Lessons from the past and applications to the future. *Curr Opin Endocrinol Diabetes*. 2016;23(2):124-130.
22. Burgess S, Freitag DF, Khan H, Gorman DN, Thompson SG. Using multivariable Mendelian randomization to disentangle the causal effects of lipid fractions. *PLoS One*. 2014;9(10):e108891.
23. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45(11):1274-1283.
24. Do R, Willer CJ, Schmidt EM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet*. 2013;45(11):1345-1352.
25. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47(10):1121-1130.
26. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 1972;18(6):499-502.
27. Di Angelantonio E, Sarwar N, Perry P, et al. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA*. 2009;302(18):1993-2000.
28. Fritz MS, Mackinnon DP. Required sample size to detect the mediated effect. *Psychol Sci*. 2007;18(3):233-239.
29. White J, Sofat R, Hemani G, et al. Plasma urate concentration and risk of coronary heart disease: A Mendelian randomisation analysis. *Lancet Diabetes Endocrinol*. 2016;4(4):327-336.
30. Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: How to interpret evidence of shared genetic predictors. *J Clin Epidemiol*. 2016;69:208-216.
31. Relton CL, Davey Smith G. Is epidemiology ready for epigenetics? *Int J Epidemiol*. 2012;41(1):5-9.
32. Suhre K, Raffler J, Kastenmuller G. Biochemical insights from population studies with genetics and metabolomics. *Arch Biochem Biophys*. 2016;589:168-176.
33. Kastenmuller G, Raffler J, Gieger C, Suhre K. Genetics of human metabolism: An update. *Hum Mol Genet*. 2015;24(R1):R93-R101.
34. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: The role of the I2 statistic. *Int J Epidemiol*. 2016;45(6):1961-1974.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Statistics in Medicine*. 2017;36:4705–4718. https://doi.org/10.1002/sim.7492

Appendix E

Appendix to paper 2

Appendix to the published paper contained in Appendix D.

Web Appendix

A1 Sample software code

We provide sample code written in R to perform the analyses described in this paper. The associations of the genetic variants with the risk factors are denoted `bXk` with standard error `bXkse`, where `k = 1, ..., K`. The associations of the genetic variants with the outcome are denoted `bY` with standard error `bYse`. The code for the multivariable models will be based on three risk factors and can be easily adapted to include the appropriate number of risk factors. It will be assumed that the causal effect of risk factor 1 on the outcome is of primary interest and all the genetic variants are uncorrelated.

Inverse-variance weighted estimate:

The inverse-variance weighted (IVW) estimate using summary statistics (equation 2) can be calculated by:

```
thetaUI     = sum(bY*bX1*bYse^-2)/sum(bX1^2*bYse^-2)
se_thetaUI  = 1/sqrt(sum(bX1^2*bYse^-2))
```

The same IVW estimate using summary statistics can be obtained using weighted linear regression (equation 3):

```
thetaUI            = summary(lm(bY~bX1-1, weights=bYse^-2))$coef[1]
se_thetaUI.fixed   = summary(lm(bY~bX1-1, weights=bYse^-2))$coef[1,2]/
                          summary(lm(bY~bX1-1, weights=bYse^-2))$sigma
se_thetaUI.random  = summary(lm(bY~bX1-1, weights=bYse^-2))$coef[1,2]/
                          min(summary(lm(bY~bX1-1, weights=bYse^-2))$sigma,1)
```

In the fixed-effect model we divide the standard error of the causal estimate by the estimated residual standard error to force the residual standard error to be 1. For the multiplicative random-effect model the standard error is divided by the estimated residual standard error when the variability in the genetic associations is less than expected by chance (underdispersion). When there is evidence of heterogeneity between the causal estimates (overdispersion) the standard error is unaltered. The multiplicative random-effects model will result in a larger standard error compared to the fixed-effect model if there is heterogeneity between the causal estimates. The causal estimate obtained from the fixed- and multiplicative random-effects models will be the same.

Univariable MR-Egger:

The univariable MR-Egger method is the same as the IVW method using weighted linear regression except the intercept term is estimated rather than being set to zero. Testing whether the intercept term is equal to zero is equivalent to testing for directional pleiotropy and the validity of the InSIDE assumption. The genetic associations with the risk factor `bX1` and outcome `bY` must be orientated with respect to the risk increasing or decreasing allele of the risk factor. Under the MR-Egger model, multiplicative random-effects should be used as the presence of pleiotropy will lead to overdispersion. Since the residual standard error is estimated, we use the t-distribution with $J-2$ degrees of freedom for inference.

```
#Orientation of the genetic associations
bY<-ifelse(bX1>0, bY, bY*-1)
bX1<-abs(bX1)
#Causal estimate
thetaUE            = summary(lm(bY~bX1, weights=bYse^-2))$coef[2]
se_thetaUE.random  = summary(lm(bY~bX1, weights=bYse^-2))$coef[2,2]/
                          min(summary(lm(bY~bX1, weights=bYse^-2))$sigma,1)
lb_thetaUE         = thetaUE - qt(0.975,df=length(bX1)-2)*se_thetaUE.random
ub_thetaUE         = thetaUE + qt(0.975,df=length(bX1)-2)*se_thetaUE.random
p_thetaUE          = 2*(1-pt(abs(thetaUE/se_thetaUE.random),df=length(bX1)-2))
#Test for directional pleiotropy
interUE            = summary(lm(bY~bX1, weights=bYse^-2))$coef[1]
se_interUE.random  = summary(lm(bY~bX1, weights=bYse^-2))$coef[1,2]/
                          min(summary(lm(bY~bX1, weights=bYse^-2))$sigma,1)
p_interUE          = 2*(1-pt(abs(interUE/se_interUE.random),df=length(bX1)-2))
```

Multivariable IVW:

The multivariable IVW method expands the IVW method using weighted linear regression by estimating the causal effects of the additional risk factors on the outcome. We will include additional two risk factors and assume the causal estimate of interest is the effect of risk factor 1 on the outcome. Either fixed- or multiplicative random-effects can be used to estimate the standard error of the causal effect.

```
theta1MI           = summary(lm(bY~bX1+bX2+bX3-1, weights=bYse^-2))$coef[1]
se_theta1MI.fixed  = summary(lm(bY~bX1+bX2+bX3-1, weights=bYse^-2))$coef[1,2]/
                           summary(lm(bY~bX1+bX2+bX3-1, weights=bYse^-2))$sigma
se_theta1MI.random = summary(lm(bY~bX1+bX2+bX3-1, weights=bYse^-2))$coef[1,2]/
                           min(summary(lm(bY~bX1+bX2+bX3-1, weights=bYse^-2))$sigma,1)
```

Multivariable MR-Egger:

The multivariable MR-Egger method is equivalent to the multivariable IVW method using weighted linear regression except the intercept is estimated rather than being set to zero. Testing whether the intercept term is equal to zero is equivalent to testing

for directional pleiotropy and the validity of the InSIDE assumption. As with univariable MR-Egger, the standard errors should be calculated from the multiplicative random-effects model. The genetic associations should be orientated with respect to the risk increasing or decreasing allele of the risk factor of interest. In this sample code we will assume the causal effect of risk factor 1 is of primary interest. Since the residual standard error is estimated for the multivariable MR-Egger model we use the t-distribution with $J - (K + 1)$ degrees of freedom for inference.

```
#Orientation of the genetic associations with respect to X1
clist<-c("bX2","bX3","bY")
for (var in clist){
  eval(parse(text=paste0(var,"<-ifelse(bX1>0,",var,",",var,"*-1)")))
}
bX1<-abs(bX1)
#Causal estimate for X1
theta1ME            = summary(lm(bY~bX1+bX2+bX3, weights=bYse^-2))$coef[2]
se_theta1ME.random  = summary(lm(bY~bX1+bX2+bX3, weights=bYse^-2))$coef[2,2]/
                          min(summary(lm(bY~bX1+bX2+bX3, weights=bYse^-2))$sigma,1)
lb_theta1ME         = theta1ME - qt(0.975,df=length(bX1)-4)*se_theta1ME.random
ub_theta1ME         = theta1ME + qt(0.975,df=length(bX1)-4)*se_theta1ME.random
p_theta1ME          = 2*(1-pt(abs(theta1ME/se_theta1ME.random),df=length(bX1)-4))
#Test for directional pleiotropy
interME            = summary(lm(bY~bX1+bX2+bX3, weights=bYse^-2))$coef[1]
se_interME.random  = summary(lm(bY~bX1+bX2+bX3, weights=bYse^-2))$coef[1,2]/
                          min(summary(lm(bY~bX1+bX2+bX3, weights=bYse^-2))$sigma,1)
p_interME          = 2*(1-pt(abs(interME/se_interME.random),df=length(bX1)-4))
```

A2 Comparison between the precision of the causal estimates from univariable and multivariable MR-Egger

In this section, we compare the precision of the causal estimates from the univariable ($\hat{\theta}_{1UE}$) and multivariable ($\hat{\theta}_{1ME}$) MR-Egger models. For the multivariable model, we consider the genetic associations $\boldsymbol{\beta_{X_k}}$ with two risk factors ($k = 2$), where the variance of the multivariable MR-Egger estimate $\hat{\theta}_{1ME}$ is given by:

$$\begin{aligned}\text{var}(\hat{\theta}_{1ME}) &= \frac{\phi^2 \,\text{var}(\boldsymbol{\beta_{X_2}})}{N(\text{var}(\boldsymbol{\beta_{X_1}})\,\text{var}(\boldsymbol{\beta_{X_2}}) - \text{cov}(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})^2)} \\ &\propto [\text{var}(\boldsymbol{\beta_{X_1}})(1 - \text{cor}(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})^2)]^{-1} \end{aligned} \tag{1}$$

Under a fixed-effect model, the variance of the univariable MR-Egger estimate is proportional to the inverse of $\text{var}(\boldsymbol{\beta_{X_1}})$.[1] The estimate from the multivariable MR-Egger model $\hat{\theta}_{1ME}$ will be more precise than its univariable counterpart $\hat{\theta}_{1UE}$ if:

$$\frac{1}{\text{var}(\boldsymbol{\beta_{X_1}})} > \frac{1}{\text{var}(\boldsymbol{\beta_{X_1}})(1 - \text{cor}(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})^2)} \tag{2}$$

From the above inequality, $\hat{\theta}_{1UE}$ will always be more precise than $\hat{\theta}_{1ME}$ when $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ are correlated. Under a multiplicative random-effects model (used throughout this paper), the variance of the residual error is estimated under the univariable MR-Egger model (ϕ^2_{UE}) and the multivariable MR-Egger model (ϕ^2_{ME}). For $\hat{\theta}_{1ME}$ to be more precise than $\hat{\theta}_{1UE}$, we require:

$$\frac{\phi^2_{UE}}{\text{var}(\boldsymbol{\beta_{X_1}})} > \frac{\phi^2_{ME}}{\text{var}(\boldsymbol{\beta_{X_1}})(1 - \text{cor}(\boldsymbol{\beta_{X_1}}, \boldsymbol{\beta_{X_2}})^2)} \tag{3}$$

If $\boldsymbol{\beta_{X_2}}$ explains additional independent variability in the genetic associations with the outcome $\boldsymbol{\beta_Y}$, and $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ are independent, then the estimate from multivariable MR-Egger will be more precise than the estimate from univariable MR-Egger. If $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ are correlated, then the precision of $\hat{\theta}_{1ME}$ will depend upon the strength of the correlation between $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$, and the amount of additional independent variability $\boldsymbol{\beta_{X_2}}$ explains in $\boldsymbol{\beta_Y}$. As the correlation between $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ increases, and $\boldsymbol{\beta_{X_2}}$ explains no additional independent variability in $\boldsymbol{\beta_Y}$, the precision of the multivariable MR-Egger estimate $\hat{\theta}_{1ME}$ will decrease.

A3 Summary statistics from the simulation study

The IVW and MR-Egger methods do not account for uncertainty in the genetic associations with the risk factor, referred to by Bowden et al as NO Measurement Error (NOME).[1] If there is substantial uncertainty in these association estimates and in a two-sample setting, the causal effect estimate from univariable MR-Egger may be biased towards the null. Bowden et al have shown that the relative attenuation in the MR-Egger estimate is approximately equal to the I^2 statistic from the meta-analysis of the weighted associations with the exposure $\hat{\beta}_{Xj}\,\mathrm{se}(\hat{\beta}_{Yj})^{-1}$, with standard errors $\mathrm{se}(\hat{\beta}_{Xj})\,\mathrm{se}(\hat{\beta}_{Yj})^{-1}$.[1] The I^2 statistic lies between 0 and 1, with smaller values corresponding to more biased MR-Egger estimates. If the I^2 statistic is close to 1, then there should be little or no attenuation of the causal estimate from the univariable MR-Egger method. Bowden et al recommend that methods to account for this uncertainty be considered if the I^2 statistic is less than 90%.[1]

The F-statistic is often reported in Mendelian randomization studies as a measurement of the strength of the instrumental variables, with larger values representing stronger instruments. For a two-sample Mendelian randomization analysis with summarized data, the F-statistic for each genetic variant j can be approximated by $F_j = \hat{\beta}_{X_j}^2 / \mathrm{se}(\hat{\beta}_{X_j})^2$. We use this approximation below.

The data-generating model used in the simulation study did not provide the standard errors of the genetic associations with the three risk factors $\mathrm{se}(\hat{\boldsymbol{\beta}}_{\boldsymbol{X_k}})$, as they were not required for the methods considered. To estimate the mean values of the F-statistics and I^2 statistics, we must make assumptions about the values of these standard errors. We assume that the genetic associations with the risk factors are provided on the standard deviation scale. If the associations were estimated from a sample size of 10 000, this results in a standard error of 0.01. Assuming that the standard errors of the genetic associations with the three risk factors are 0.01 across the 185 genetic variants, we obtain the mean F-statistics and I^2 statistics displayed in Table A1 and Table A2. The I^2 statistics (reported as a %) are close to 100% across the different scenarios. These results are consistent with the simulation study where the causal estimates from the univariable and multivariable MR-Egger methods showed no attenuation towards the null.

Table A1: Mean F-statistic and I^2 statistic (reported as a %) for a null ($\theta_1 = 0$) and positive ($\theta_1 = 0.3$) causal effect where $\boldsymbol{\beta_{X_k}}$ are generated independently for all k.

	$\hat{\beta}_{X_{1j}}$		$\hat{\beta}_{X_{2j}}$		$\hat{\beta}_{X_{3j}}$	
	F-statistic	I^2 statistic	F-statistic	I^2 statistic	F-statistic	I^2 statistic
Null causal effect: $\boldsymbol{\theta_1 = 0}$						
1. No pleiotropy, InSIDE satisfied						
	363.3	99.5	208.8	99.2	425.7	99.6
2. Balanced pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0,0.004)$	364.3	99.5	209.1	99.2	425.6	99.6
3. Directional pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	364.4	99.5	208.9	99.2	425.6	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	363.5	99.5	209.6	99.2	424.9	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	364.0	99.5	209.2	99.2	425.5	99.6
4. Directional pleiotropy, InSIDE violated						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	364.1	99.5	208.8	99.2	425.4	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.4	99.5	208.7	99.2	425.4	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	363.8	99.5	209.1	99.2	425.1	99.6
Positive causal effect: $\boldsymbol{\theta_1 = 0.3}$						
1. No pleiotropy, InSIDE satisfied						
	363.9	99.5	209.2	99.2	424.7	99.6
2. Balanced pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0,0.004)$	363.7	99.5	209.1	99.2	425.0	99.6
3. Directional pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	364.1	99.5	209.0	99.2	425.2	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.3	99.5	208.6	99.2	425.5	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	363.8	99.5	209.1	99.2	424.7	99.6
4. Directional pleiotropy, InSIDE violated						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	363.6	99.5	209.1	99.2	424.8	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.6	99.5	208.9	99.2	424.8	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	364.0	99.5	208.9	99.2	425.9	99.6

Table A2: Mean F-statistic and I^2 statistic (reported as a %) for a null ($\theta_1 = 0$) and positive ($\theta_1 = 0.3$) causal effect with $\boldsymbol{\beta_{X_k}}$ being correlated for all k.

	$\hat{\beta}_{X_{1j}}$		$\hat{\beta}_{X_{2j}}$		$\hat{\beta}_{X_{3j}}$	
	F-statistic	I^2 statistic	F-statistic	I^2 statistic	F-statistic	I^2 statistic
Null causal effect: $\boldsymbol{\theta_1 = 0}$						
1. No pleiotropy, InSIDE satisfied						
	364.1	99.5	208.5	99.2	424.5	99.6
2. Balanced pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0,0.004)$	363.8	99.5	209.4	99.2	424.4	99.6
3. Directional pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	363.5	99.5	208.8	99.2	424.6	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.3	99.5	209.0	99.2	425.1	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	364.3	99.5	208.9	99.2	424.7	99.6
4. Directional pleiotropy, InSIDE violated						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	364.0	99.5	208.9	99.2	425.0	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.0	99.5	209.4	99.2	425.2	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	364.3	99.5	209.1	99.2	425.2	99.6
Positive causal effect: $\boldsymbol{\theta_1 = 0.3}$						
1. No pleiotropy, InSIDE satisfied						
	364.0	99.5	208.9	99.2	425.2	99.6
2. Balanced pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0,0.004)$	364.0	99.5	208.8	99.2	425.1	99.6
3. Directional pleiotropy, InSIDE satisfied						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	363.5	99.5	209.1	99.2	425.5	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	363.9	99.5	209.1	99.2	424.6	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	364.0	99.5	209.1	99.2	425.8	99.6
4. Directional pleiotropy, InSIDE violated						
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	364.1	99.5	208.8	99.2	425.3	99.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	364.7	99.5	208.8	99.2	425.4	99.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	363.7	99.5	208.9	99.2	424.5	99.6

A4 Details and results from the simulation study investigating causal relationships between risk factors

To investigate the behaviour of the multivariable MR-Egger method when causal relationships between risk factors exist, additional simulations were performed where X_2 was causally dependent on X_1. We assume that X_2 is causally dependent on X_1, and the causal effect of X_1 on X_2 is γ. The total causal effect of X_1 on Y is $\theta_1 + \gamma\theta_2$; consisting of the direct effect (θ_1) and the indirect effect via X_2 ($\gamma\theta_2$). The simulations outlined in Section 4 were repeated with the second line in the data generating model replaced with:

$$\beta_{Y_j} = \alpha'_j + \theta_1|\beta_{X_{1j}}| + \theta_2(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|) + \theta_3\beta_{X_{3j}} + \epsilon_j \tag{4}$$

The causal effect of X_1 on X_2 (γ) was set to 0.5. All other parameters were taken as in the original simulation study. $|\beta_{X_{1j}}|$, $(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|)$, and $\beta_{X_{3j}}$ were the covariates included in the multivariable IVW and multivariable MR-Egger models. Note that the functional relationship between X_1 and X_2 induces a correlation structure between the covariates $|\beta_{X_{1j}}|$ and $(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|)$ included in the multivariable models, even when $\boldsymbol{\beta_{X_1}}$ and $\boldsymbol{\beta_{X_2}}$ are generated independently. To account for the additional uncertainty in β_{Y_j}, the weights for univariable MR-Egger are given by equation 5, while the weights for multivariable IVW and multivariable MR-Egger were the same as the original simulation study (equation 15).

$$\mathrm{se}(\beta_{Y_j})^{-2} = ({\epsilon_j}^2 + {\sigma_{\alpha'}}^2 + {\theta_2}^2{\sigma_2}^2 + (\theta_2\gamma)^2{\sigma_1}^2 + 2\theta_2\gamma\rho_{12}\sigma_1\sigma_2 + {\theta_3}^2{\sigma_3}^2)^{-1} \tag{5}$$

Results

The results from the simulations that included a causal relationship between X_1 and X_2, using 10 000 simulated datasets, are presented in Web Table A3 ($\boldsymbol{\beta_{X_k}}$ generated independently, with the functional relationship between X_1 and X_2 inducing a correlation structure between $|\beta_{X_{1j}}|$ and $(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|)$) and Web Table A4 ($\boldsymbol{\beta_{X_k}}$ correlated).

$\boldsymbol{\beta_{X_k}}$ generated independently, with a correlation structure between the covariates $|\beta_{X_{1j}}|$ and $(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|)$: In scenarios where there was no bias in the original set of simulations, the multivariable IVW and multivariable MR-Egger methods consistently estimated the direct effect of X_1 on Y (θ_1), whilst the univariable MR-Egger method consistently estimated the total causal effect of X_1 on Y ($\theta_1 + \gamma\theta_2$).

Bias for the multivariable IVW method was present in scenarios 3 and 4 only, as in the original simulation study (Tables 3 and 4). Compared to the results in Table 3, precision and power to detect a causal effect were reduced for the multivariable IVW and multivariable MR-Egger methods. This reduction in power may be due to the correlation structure between $|\beta_{X_{1j}}|$ and $(\beta_{X_{2j}} + \gamma|\beta_{X_{1j}}|)$, and the multivariable models conditioning on a mediator. Univariable and multivariable MR-Egger methods produced biased estimates of the total and direct causal effects in scenario 4 (InSIDE violated) only. Unlike the original simulation study, precision and power to detect a causal effect were always better for the univariable MR-Egger method.

$\boldsymbol{\beta_{X_k}}$ correlated: The multivariable IVW and multivariable MR-Egger methods estimated the direct effect of X_1 on Y, as in the independently generated setting. As with the original simulations (Tables 3 and 4), the InSIDE assumption for univariable MR-Egger was violated for all four scenarios, resulting in biased point estimates. However, as with the original simulation study, the multivariable InSIDE assumption was satisfied for scenarios 1,2 and 3, and so causal estimates from multivariable MR-Egger were unbiased. There was a more noticeable reduction in the precision and power to detect a causal effect for the multivariable IVW and multivariable MR-Egger methods under the correlated setting.

Table A3: Performance of multivariable IVW, univariable MR-Egger and multivariable MR-Egger with respect to $\hat{\theta}_1$ for a null ($\theta_1 = 0$) and positive ($\theta_1 = 0.3$) causal effect where $\boldsymbol{\beta_{X_k}}$ are generated independently for all k (with a correlation structure between the covariates $|\beta_{X_{1j}}|$ and $(\beta_{X_{2j}}+\gamma|\beta_{X_{1j}}|)$), with a causal effect of X_1 on X_2 ($\gamma = 0.5$). All tests were performed at the 5% level of significance.

	Multivariable IVW		**Univariable MR-Egger**			**Multivariable MR-Egger**		
	Mean $\hat{\theta}_1$ (mean SE)	Power, %	Mean $\hat{\theta}_1$ (mean SE)	Power, % Intercept	Causal	Mean $\hat{\theta}_1$ (mean SE)	Power, % Intercept	Causal
Null causal effect: $\boldsymbol{\theta_1 = 0}$								
1. No pleiotropy, InSIDE satisfied								
	0.000 (0.057)	3.5	0.051 (0.158)	8.9	5.8	0.001 (0.090)	4.5	4.2
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	0.001 (0.127)	4.4	0.049 (0.187)	7.6	5.6	0.001 (0.178)	4.6	4.2
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.041 (0.127)	6.0	0.049 (0.187)	12.3	5.4	0.000 (0.178)	5.8	4.8
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.195 (0.128)	34.4	0.048 (0.187)	50.1	5.3	-0.001 (0.178)	36.6	4.6
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.393 (0.130)	82.3	0.052 (0.187)	91.4	5.6	0.002 (0.178)	88.4	4.7
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.076 (0.127)	9.8	0.138 (0.187)	6.4	11.6	0.088 (0.178)	4.3	7.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.231 (0.127)	45.2	0.137 (0.187)	34.4	11.9	0.088 (0.178)	21.7	8.2
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.426 (0.129)	88.3	0.141 (0.187)	83.7	11.9	0.089 (0.178)	78.2	8.1
Positive causal effect: $\boldsymbol{\theta_1 = 0.3}$								
1. No pleiotropy, InSIDE satisfied								
	0.301 (0.057)	96.3	0.353 (0.158)	9.3	62.3	0.301 (0.090)	3.9	84.6
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	0.298 (0.127)	65.4	0.350 (0.187)	7.4	47.8	0.298 (0.178)	4.4	41.2
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.338 (0.127)	75.5	0.352 (0.187)	11.8	48.3	0.300 (0.178)	6.1	41.1
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.494 (0.128)	95.2	0.348 (0.188)	49.2	46.9	0.298 (0.179)	36.8	40.3
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.689 (0.130)	99.6	0.347 (0.188)	91.5	47.1	0.296 (0.178)	88.2	39.6
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.375 (0.127)	82.6	0.440 (0.187)	6.6	65.7	0.390 (0.178)	4.7	60.1
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.530 (0.128)	97.0	0.438 (0.187)	34.7	65.5	0.386 (0.178)	21.7	59.9
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.728 (0.129)	99.7	0.441 (0.187)	83.6	65.8	0.390 (0.178)	78.5	60.1

Abbreviations: MR, Mendelian randomization; SE, standard error; IVW, inverse-variance weighted; InSIDE, Instrument Strength Independent of Direct Effect.

Table A4: Performance of multivariable IVW, univariable MR-Egger and multivariable MR-Egger with $\boldsymbol{\beta_{X_k}}$ being correlated for all k, and a causal effect of X_1 on X_2

	Multivariable IVW		**Univariable MR-Egger**			**Multivariable MR-Egger**		
	Mean $\hat{\theta}_1$ (mean SE)	Power, %	Mean $\hat{\theta}_1$ (mean SE)	Power, % Intercept	Causal	Mean $\hat{\theta}_1$ (mean SE)	Power, % Intercept	Causal
Null causal effect: $\boldsymbol{\theta_1 = 0}$								
1. No pleiotropy, InSIDE satisfied								
	0.000 (0.062)	4.1	0.146 (0.158)	3.9	15.6	0.000 (0.097)	4.0	4.0
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	0.000 (0.137)	4.5	0.146 (0.188)	4.1	11.9	0.000 (0.190)	4.6	4.7
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.041 (0.137)	5.7	0.151 (0.187)	5.4	12.8	0.003 (0.189)	5.7	4.4
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.209 (0.138)	34.2	0.148 (0.187)	32.8	12.6	0.000 (0.190)	36.9	4.7
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.422 (0.140)	82.2	0.151 (0.188)	83.0	12.9	0.004 (0.190)	89.0	4.8
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.053 (0.137)	6.2	0.235 (0.188)	4.3	25.7	0.069 (0.189)	4.9	6.4
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.218 (0.137)	37.2	0.235 (0.188)	20.3	26.4	0.067 (0.189)	21.8	6.7
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.429 (0.139)	84.3	0.238 (0.188)	71.3	26.7	0.071 (0.189)	79.2	6.6
Positive causal effect: $\boldsymbol{\theta_1 = 0.3}$								
1. No pleiotropy, InSIDE satisfied								
	0.299 (0.062)	94.7	0.446 (0.158)	4.1	79.7	0.300 (0.096)	4.0	81.3
2. Balanced pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0,0.004)$	0.301 (0.137)	60.5	0.445 (0.187)	4.5	66.6	0.300 (0.189)	4.6	37.0
3. Directional pleiotropy, InSIDE satisfied								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.339 (0.137)	69.9	0.443 (0.188)	5.7	66.1	0.296 (0.190)	6.0	36.1
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.510 (0.138)	94.2	0.449 (0.188)	32.6	67.7	0.302 (0.190)	37.3	37.2
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.715 (0.140)	99.2	0.445 (0.187)	83.4	66.9	0.298 (0.189)	89.4	36.8
4. Directional pleiotropy, InSIDE violated								
$\alpha'_j \sim \mathcal{N}(0.01,0.004)$	0.353 (0.137)	73.1	0.534 (0.188)	4.4	79.4	0.367 (0.189)	4.6	50.6
$\alpha'_j \sim \mathcal{N}(0.05,0.004)$	0.519 (0.138)	95.1	0.534 (0.188)	20.3	79.6	0.366 (0.190)	21.7	50.5
$\alpha'_j \sim \mathcal{N}(0.1,0.004)$	0.728 (0.139)	99.5	0.533 (0.188)	72.5	79.6	0.368 (0.189)	80.1	51.0

Abbreviations: MR, Mendelian randomization; SE, standard error; IVW, inverse-variance weighted; InSIDE, Instrument Strength Independent of Direct Effect.

A5 Correlated genetic variants

The methods discussed in this article have assumed that the genetic variants are uncorrelated (not in linkage disequilibrium). There may, however, be cases where using multiple correlated variants from the same gene region will be more efficient than using uncorrelated variants from different gene regions.[2] If the genetic variants are in partial linkage disequilibrium, and each variant explains independent variation in the risk factor, then the inclusion of these variants will increase the power of the MR study. The precision of a MR study will not increase, however, if the variants are perfectly correlated.

If correlated variants are included in an MR study, using summarized level data, the analysis should account for the correlation structure of the variants. If the correlation of the variants is not taken into consideration, the causal estimate will be too precise and this may lead to inappropriate inferences. To account for the correlation between the genetic variants for the univariable and multivariable IVW methods, we can use generalized weighted linear regression of the genetic associations, where the correlations of the variants are included in the weighting matrix, with the intercept set to zero.[2,3]

If $\Omega_{st} = \text{se}(\hat{\beta}_{Y_s})\,\text{se}(\hat{\beta}_{Y_t})\rho_{st}$, where ρ_{st} is the correlation between variants s and t, then the causal estimate from a weighted generalised linear regression for univariable MR is:

$$\hat{\theta}_{UIC} = (\hat{\beta}_{X_j}^T \Omega^{-1} \hat{\beta}_{X_j})^{-1} \hat{\beta}_{X_j}^T \Omega^{-1} \hat{\beta}_{Y_j} \tag{6}$$

with the standard error of the causal estimate:

$$\hat{\theta}_{UIC} = \sqrt{(\hat{\beta}_{X_j}^T \Omega^{-1} \hat{\beta}_{X_j})^{-1}} \tag{7}$$

Whilst the univariable MR-Egger estimates can be obtained by fitting the same generalized weighted linear regression model, but allowing the intercept term to be estimated, the effect of using correlated genetic variants in the univariable MR-Egger method has not been considered in detail. Further investigation into the impact correlated variants may have on the interpretation of the direct effect, and the InSIDE assumption, must be considered at the univariable level first, and then expanded to multivariable MR-Egger.

References

1. Bowden J, Del Greco M. F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol.* 2016;45(6):1961–1974.

2. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med.* 2016;35(11):1880–1906.

3. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol.* 2015;181(4):251–260.

Appendix F

Paper 3: Factorial Mendelian randomization: using genetic variants to assess interactions

Published paper based on the work in Chapter 5.

International Journal of Epidemiology, 2019, 1–12
doi: 10.1093/ije/dyz161
Original article

Original article

Factorial Mendelian randomization: using genetic variants to assess interactions

Jessica M B Rees,[1,2] Christopher N Foley,[3] and Stephen Burgess [1,3]*

[1]Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, [2]Edinburgh Clinical Trials Unit, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK and [3]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

*Corresponding author. MRC Biostatistics Unit, Cambridge Institute of Public Health, Robinson Way, Cambridge CB2 0SR, UK. E-mail: sb452@medschl.cam.ac.uk

Editorial decision 21 June 2019; Accepted 9 July 2019

Abstract

Background: Factorial Mendelian randomization is the use of genetic variants to answer questions about interactions. Although the approach has been used in applied investigations, little methodological advice is available on how to design or perform a factorial Mendelian randomization analysis. Previous analyses have employed a 2×2 approach, using dichotomized genetic scores to divide the population into four subgroups as in a factorial randomized trial.

Methods: We describe two distinct contexts for factorial Mendelian randomization: investigating interactions between risk factors, and investigating interactions between pharmacological interventions on risk factors. We propose two-stage least squares methods using all available genetic variants and their interactions as instrumental variables, and using continuous genetic scores as instrumental variables rather than dichotomized scores. We illustrate our methods using data from UK Biobank to investigate the interaction between body mass index and alcohol consumption on systolic blood pressure.

Results: Simulated and real data show that efficiency is maximized using the full set of interactions between genetic variants as instruments. In the applied example, between 4- and 10-fold improvement in efficiency is demonstrated over the 2×2 approach. Analyses using continuous genetic scores are more efficient than those using dichotomized scores. Efficiency is improved by finding genetic variants that divide the population at a natural break in the distribution of the risk factor, or else divide the population into more equal-sized groups.

Conclusions: Previous factorial Mendelian randomization analyses may have been underpowered. Efficiency can be improved by using all genetic variants and their interactions as instrumental variables, rather than the 2×2 approach.

Key words: Mendelian randomization, instrumental variables, interaction, causal inference, factorial randomized trial

Key Messages

- Factorial Mendelian randomization is an extension of the Mendelian randomization paradigm to answer questions about interactions.
- There are two contexts in which factorial Mendelian randomization can be used: for investigating interactions between risk factors, and interactions between pharmacological interventions on risk factors.
- While most applications of factorial Mendelian randomization have dichotomized the population as in a 2×2 factorial randomized trial, this approach is generally inefficient for detecting statistical interactions.
- In the first context, efficiency is maximized by including all genetic variants and their cross-terms as instrumental variables for the two risk factors and their product term.
- In the second context, efficiency is maximized by using continuous genetic scores rather than dichotomized scores.

Introduction

Mendelian randomization is the use of genetic variants as proxies for interventions on risk factors to answer questions of cause and effect using observational data.[1,2] Formally, Mendelian randomization can be viewed as instrumental variable (IV) analysis using genetic variants as IVs.[3,4] Factorial Mendelian randomization is the use of genetic variants to answer questions about interactions. It does this by proposing a statistical model for the outcome as a function of the risk factors (or their genetic predictors) and a product term.

A statistical interaction is observed when the coefficient for the product term in the model is non-zero. A statistical interaction in the causal model for the outcome may represent a causal interaction, meaning that the effect of one risk factor on the outcome is dependent upon the value of the other risk factor.[5,6] This may arise due to a functional or biological interaction, in which there is a mechanistic connection between the two risk factors in how they influence the outcome. However, a statistical interaction may also arise due to non-linearity in the effect of a risk factor, or due to effect modification, in which the effect of one risk factor varies in strata of the other. Hereafter, we take the word 'interaction' to mean a statistical interaction in the causal model for the outcome, without implying a functional interaction between the risk factors.

Factorial Mendelian randomization was proposed in the seminal paper on Mendelian randomization by Davey Smith and Ebrahim in 2003.[1] The term is credited by the authors to Sheila Bird (https://en.wikipedia.org/wiki/Sheila_Bird). However, the idea was not readily taken up in applied practice. The concept was raised again by Davey Smith and Hemani in a 2014 review,[7] who suggested that genetic predictors of obesity and alcohol consumption could be used to investigate the interaction between the two risk factors on risk of liver disease. This question was investigated for markers of liver function using data from the Copenhagen General Population Study in 2018;[8] no evidence for an interaction was found.

In parallel to this, the term factorial Mendelian randomization has been used for analyses employing genetic variants as proxies for pharmacological interventions. Ference *et al.* performed factorial Mendelian randomization to compare the effect of lowering low density lipoprotein (LDL) cholesterol levels on the risk of coronary heart disease (CHD) with two different LDL-cholesterol lowering agents (ezetimibe and statin), and with a combination of both.[9] Genetic variants associated with LDL-cholesterol were identified in the *NPC1L1* gene (proxies for ezetimibe), and the *HMGCR* gene (proxies for statins), and combined into separate gene scores. To mimic a 2×2 factorial randomized trial, the two gene scores were dichotomized to create a 2×2 contingency table. The gene scores were dichotomized at their median values so that the numbers of participants were balanced across the four groups. Ference has conducted similar analyses for PCSK9 inhibitors and statins,[10] and for CETP inhibitors and statins.[11] A similar 2×2 approach was used in each case, as well as in the analysis of obesity and alcohol mentioned above.[8]

In this paper, we consider various aspects relating to the conceptualization, design, analysis and interpretation of a factorial Mendelian randomization investigation. First, we demonstrate the analogy between factorial Mendelian randomization and a factorial randomized trial, we make a connection with multivariable Mendelian randomization, and we describe two contexts in which factorial Mendelian randomization may have utility: for investigating interactions between risk factors, and for investigating interactions between pharmacological interventions on risk factors. We present simulated data demonstrating that the 2×2 approach to analysis, while being conceptually appealing, is inefficient for detecting interactions. The same conclusion is reached in an applied investigation considering interactions between body mass index (BMI) and alcohol consumption on blood pressure using data from UK Biobank. Finally, we discuss the implications of our work for applied factorial Mendelian randomization investigations.

Factorial randomized trial		Randomization of A	
		Control	Treatment A
Randomization of B	Control	Incidence under usual care	Incidence under intervention in A
	Treatment B	Incidence under intervention in B	Incidence under intervention in A and B

Factorial Mendelian randomization		Genetic score 1	
		Below median	Above median
Genetic score 2	Below median	Risk factor 1 lower, risk factor 2 lower	Risk factor 1 higher, risk factor 2 lower
	Above median	Risk factor 1 lower, risk factor 2 higher	Risk factor 1 higher, risk factor 2 higher

Figure 1. Comparison of a factorial randomized clinical trial and a factorial Mendelian randomization investigation with a 2×2 approach (adapted from[15]).

Methods

Factorial randomized trials and Mendelian randomization

A factorial randomized trial allows for the simultaneous assessment of two or more treatments in a single study.[12] In its simplest form, a 2×2 factorial trial investigates the effect of two binary treatments A and B on a binary outcome *Y*. Participants are randomly allocated to one of four groups: to receive treatment A only; to receive treatment B only; to receive both treatments A and B; or to receive neither treatment A nor B. The analogy between Mendelian randomization and a randomized trial has been made many times,[13,14] and the analogy between factorial Mendelian randomization and a factorial randomized trial has also been made previously in the context of multivariable Mendelian randomization (Figure 1, adapted from[15]).

Multivariable Mendelian randomization was motivated by the problem that some genetic variants are associated with multiple risk factors, such that it is impossible to find genetic variants that are specifically associated with a particular risk factor.[15] For illustration, we assume there are two risk factors (X_1 and X_2), and fit a model for the outcome in terms of the risk factors:

$$E(Y|X_1, X_2) = \theta_0 + \theta_1 X_1 + \theta_2 X_2. \quad (1)$$

We assume that we have genetic variants *G* that satisfy the multivariable IV assumptions for risk factors X_1 and X_2.[15] Specifically:

i. Each variant is associated with at least one of the risk factors.
ii. Each risk factor is associated with at least one of the genetic variants.
iii. Variants are not confounded in their associations with the outcome.
iv. Variants are not associated with the outcome conditional on the risk factors and confounders.

If we have at least two genetic variants that are valid multivariable IVs for X_1 and X_2, then causal effects θ_1 and θ_2 can be estimated from the two-stage least squares method by first regressing the risk factors on the genetic variants, and then regressing the outcome on the fitted values of the risk factors from the first-stage regressions.[16] If summarized data on the genetic associations with the outcome ($\hat{\beta}_Y$) and the risk factors ($\hat{\beta}_{X1}$, $\hat{\beta}_{X2}$) are available, then the same estimates can be obtained by weighted linear regression of the beta-coefficients with the intercept set to zero:

$$E(\hat{\beta}_Y|\hat{\beta}_{X1}, \hat{\beta}_{X2}) = \theta_1 \hat{\beta}_{X1} + \theta_2 \hat{\beta}_{X2}, \quad (2)$$

where weights are the reciprocals of variances of the gene–outcome associations $\mathrm{se}(\hat{\beta}_Y)^{-2}$.[17]

In the language of a factorial randomized trial, this is referred to as an analysis performed 'at the margins'.[18] Estimates represent the average direct effect of each of the risk factors.[19] If there is an interaction between the risk factors, then these are marginal estimates—they are averaged over the distribution of the other risk factor.

We can extend multivariable Mendelian randomization by adding a term to the outcome model to estimate an interaction between the risk factors:

$$E(Y|X_1, X_2) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_{12} X_{12} \quad (3)$$

where X_{12} is the product $X_1 \times X_2$, and θ_{12} is the interaction effect on an additive scale. In a factorial randomized trial, this is referred to as an analysis performed 'inside the table', as in a 2×2 setting, the interaction can be estimated from the 2×2 contingency table.[20] A factorial Mendelian randomization analysis is primarily interested in assessing the presence of, and estimating the interaction effect θ_{12}.

For simplicity, we initially assume that the associations of the genetic variants with the risk factors are homogeneous in the population and do not vary with time, also that the model relating the risk factors to the outcome is correctly specified, and the effects of the risk factors (and their product) on the outcome are also homogeneous in the population and do not vary with time. We return to the question of how to interpret estimates in this and in more realistic scenarios in the Discussion.

Two contexts: interactions between risk factors and interactions between interventions

Factorial Mendelian randomization study has been considered in two broad scenarios: (a) to estimate interaction effects between risk factors by using genetic variants as predictors of the risk factors; and (b) to identify interactions between interventions by using genetic variants as proxies for specific treatments. In the first case, the aim is to identify an interaction in the effect of two distinct risk factors on the outcome. In the second case, there may not even be two distinct risk factors (as in the example of two LDL-cholesterol lowering interventions discussed by Ference *et al.*[9]) and the aim is to identify an interaction in the associations of the genetic variants with the outcome. In this case, an interaction is inferred between the interventions for which the genetic variants are proxies. We consider these two scenarios in turn.

Interactions between risk factors

The multivariable IV assumptions imply that there is no effect of the genetic variants on the outcome except potentially indirectly via one or both of the risk factors. We divide the genetic variants into three groups: G_1 contains variants that are associated with X_1, G_2 contains variants that are associated with X_2, and G_c contains shared variants that are associated with X_1 and X_2 (Figure 2). We can now perform two-stage least squares by first regressing the risk factors X_1, X_2, and the product X_{12} on the genetic variants, and then regressing the outcome on the fitted values of these risk factors. This analysis treats X_{12} as if it is a separate risk factor unrelated to X_1 and X_2.[21] For the second-stage regression model to be identified, at least three IVs are required, as three parameters are estimated, and all risk factors (X_1, X_2, X_{12}) must be associated with at least one IV.

If we assume that the risk factors X_1 and X_2 are linear in the genetic variants:

$$\begin{aligned}\mathbb{E}[X_1|G] &= \alpha_{01} + \sum \alpha_{1j} G_{1j} + \sum \alpha_{1cj} G_{cj} \text{ and} \\ \mathbb{E}[X_2|G] &= \alpha_{02} + \sum \alpha_{2j} G_{2j} + \sum \alpha_{2cj} G_{cj},\end{aligned} \tag{4}$$

then an interaction between the risk factors means that the statistical model for the outcome includes cross-terms between the genetic variants (such as $G_{11} \times G_{21}$).[22] This motivates the use of cross-terms between the genetic variants as separate IVs.

If all the genetic variants and their cross-terms are used as IVs, then under the homogeneity assumptions, the fitted values of the risk factors and their product term can be consistently estimated, and hence the regression model for the outcome on these fitted values (as in the two-stage least squares method) will be correctly specified. Thus the homogeneity assumptions lead to consistent estimates of the parameters in equation (3).

Simulation study 1: interactions between risk factors

To investigate the performance of methods for estimating interactions between risk factors, we conduct a simulation study. We assume there are 10 genetic variants that are associated with X_1 and 10 genetic variants that are associated with X_2, and vary the number of shared variants that are associated with both X_1 and X_2 from 0 (20 distinct genetic variants, each associated with one risk factor) to 10 (all 10 genetic variants associated with both risk factors).

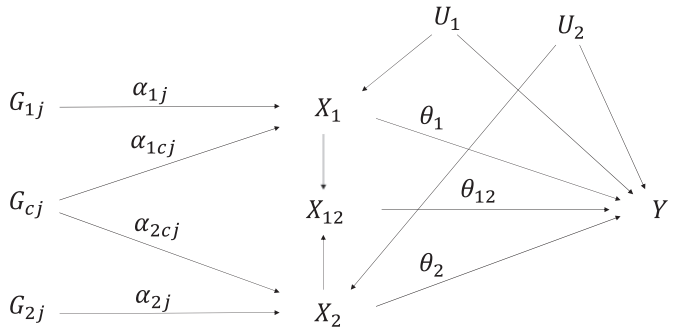


Figure 2. Causal directed acyclic graph illustrating relationships between variables in a factorial Mendelian randomization framework for two risk factors (X_1 and X_2). There are three sets of genetic variants: G_1 (affecting X_1 only), G_2 (affecting X_2 only) and G_c (shared variants, affecting X_1 and X_2). X_{12} represents the product $X_1 \times X_2$. The main effects of the risk factors X_1 and X_2 on the outcome Y are θ_1 and θ_2, and the interaction effect of X_1 and X_2 on Y is θ_{12}. U_1 and U_2 are sets of confounders.

All genetic variants are simulated as independent (i.e. not in linkage disequilibrium). We compare four methods:

Method 1. Full set of interactions: we consider as IVs all the genetic variants and all cross-terms—so when there are 3 shared variants, there are 114 IVs in total: $7+7+3=17$ linear terms, 3 quadratic terms (shared variants only), 3 shared × shared variant cross-terms, 42 shared × non-shared variant cross-terms, and 49 non-shared × non-shared variant cross-terms.

Method 2. Reduced set of interactions: we consider as IVs all the genetic variants and all cross-terms between non-shared variants—so when there are 3 shared variants, there are 17 linear terms and 49 cross-terms.

Method 3. Continuous gene scores: we construct weighted gene scores for each risk factor using external weights, and take the two gene scores and their product as IVs.

Method 4. Dichotomized gene scores: we dichotomize both gene scores at their median, and take the two dichotomized gene scores and their product as IVs. This is equivalent to a 2×2 analysis.

The data-generating model for the simulation study is provided in the Supplementary Material, available as Supplementary data at *IJE* online. Data were generated 10 000 times for each set of parameters on 10 000 individuals. Parameters were set such that the set of genetic variants explains around 10% of the variance in each risk factor. The effect of X_1 on the outcome was $\theta_1 = 0.3$, the effect of X_2 on the outcome was $\theta_2 = 0.2$, and the interaction effect of X_{12} on the outcome took values $\theta_{12} = 0.1$, 0.3, and 0.5.

Simulation study 2: interactions between interventions

We performed a further simulation study to investigate methods for detecting interactions between interventions. We assume there are 3 independent genetic variants that are proxies for intervention A, and the same for intervention B. Fewer variants are considered here as typically variants for such an analysis will come from a single gene region for each intervention.[9] We compare two approaches.

i. Continuous gene scores: we construct weighted gene scores for changes in the risk factor corresponding to each intervention using external weights, and take the two gene scores and their product as IVs.
ii. Dichotomized gene scores: we dichotomize both gene scores at their median, and take the two dichotomized gene scores and their product as IVs. This is equivalent to a 2×2 analysis.

In each case, we regressed the outcome on the IVs, and estimated an interaction term between the gene scores that act as proxies for the interventions. As before, the data-generating model for the simulation study is provided in the Supplementary Material, available as Supplementary data at *IJE* online. Data were generated 10 000 times for each set of parameters on 10 000 individuals. The interaction effect took values 0.1, 0.3, and 0.5. We varied the minor allele frequencies of the genetic variants used as proxies for interventions A and B, drawing from a uniform distribution between 0.1 and 0.2 (uncommon), or between 0.4 and 0.5 (common), and the proportion of variance explained by the genetic variants (3, 5 or 7%).

Applied example: the effects of BMI and alcohol on systolic blood pressure

Increased systolic blood pressure (SBP) is associated with a range of health conditions, including cardiovascular disease and diabetes.[23,24] Whereas there have been numerous studies highlighting the adverse effects of increased BMI on SBP,[25,26] and the adverse effects of increased alcohol consumption,[27] there has been little research on the combined effect of BMI and alcohol consumption on SBP. We illustrate factorial Mendelian randomization by performing an analysis using individual participant data from UK Biobank to estimate the interaction effect of BMI and alcohol consumption on SBP. UK Biobank is a prospective, population-based cohort consisting of ~500 000 participants aged from 40 to 69 years at baseline living in the UK. For the analysis, we considered 291 781 unrelated participants of European descent who passed data quality control measures and had genetic data available.

We used the 77 genome-wide significant variants from a meta-analysis by the Genetic Investigation of ANthropometric Traits (GIANT) consortium in participants of European ancestry to act as IVs for BMI.[28] For alcohol, we identified 10 genetic variants in the *ADH1B* gene region that have been shown to be associated with alcohol consumption.[29] We performed factorial Mendelian randomization analyses using the full set of interactions, continuous gene scores, and dichotomized gene scores. We also performed analyses separately using the lead variant from the *ADH1B* gene region (rs1229984) as the sole IV for alcohol consumption, as was done in the analysis by Carter *et al.*[8]

Results

Simulation study 1: interactions between risk factors

Results from the simulation study for estimating interactions between risk factors are displayed in Table 1 (no shared

Table 1. Simulation study results for interactions between risk factors with no shared variants: median estimate, standard deviation (SD) of estimates, median standard error (SE), empirical power (%) to reject null at 5% significance, and empirical coverage (%) of 95% confidence interval

	Median	SD	Median SE	Power (%)	Coverage (%)
Methods 1 and 2—full set of interactions:[a]					
$\theta_1 = 0.3$	0.3013	0.0917	0.0910	90.2	95.0
$\theta_2 = 0.2$	0.2022	0.0952	0.0945	57.1	94.9
$\theta_{12} = 0.1$	0.1101	0.0721	0.0718	33.7	94.6
$\theta_1 = 0.3$	0.3043	0.0918	0.0910	91.0	95.0
$\theta_2 = 0.2$	0.2034	0.0947	0.0945	57.9	95.5
$\theta_{12} = 0.3$	0.3080	0.0722	0.0718	98.8	95.2
$\theta_1 = 0.3$	0.3048	0.0911	0.0909	90.7	95.2
$\theta_2 = 0.2$	0.2050	0.0944	0.0945	58.4	95.2
$\theta_{12} = 0.5$	0.5073	0.0715	0.0718	100.0	95.2
Method 3—continuous gene scores:					
$\theta_1 = 0.3$	0.2993	0.1362	0.1333	61.4	95.4
$\theta_2 = 0.2$	0.1991	0.1415	0.1386	30.9	95.5
$\theta_{12} = 0.1$	0.1010	0.1113	0.1091	15.4	95.5
$\theta_1 = 0.3$	0.2998	0.1359	0.1332	61.9	95.6
$\theta_2 = 0.2$	0.2019	0.1405	0.1387	31.5	95.8
$\theta_{12} = 0.3$	0.3000	0.1106	0.1091	77.5	95.8
$\theta_1 = 0.3$	0.3004	0.1352	0.1331	61.5	95.4
$\theta_2 = 0.2$	0.2008	0.1409	0.1385	30.7	95.6
$\theta_{12} = 0.5$	0.4995	0.1107	0.1092	98.7	95.6
Method 4—dichotomized gene scores:					
$\theta_1 = 0.3$	0.2986	0.2155	0.2072	31.0	95.7
$\theta_2 = 0.2$	0.1989	0.2246	0.2168	15.0	96.2
$\theta_{12} = 0.1$	0.1022	0.1786	0.1720	8.0	95.9
$\theta_1 = 0.3$	0.3039	0.2145	0.2074	32.1	95.8
$\theta_2 = 0.2$	0.2047	0.2236	0.2164	15.2	96.2
$\theta_{12} = 0.3$	0.2972	0.1777	0.1722	41.8	96.0
$\theta_1 = 0.3$	0.3010	0.2148	0.2073	31.4	96.2
$\theta_2 = 0.2$	0.2002	0.2233	0.2163	15.3	96.1
$\theta_{12} = 0.5$	0.5002	0.1776	0.1718	80.7	96.1

[a]As there are no shared variants, methods 1 and 2 are equivalent.

Table 2. Simulation study results for interaction term between risk factors varying number of shared variants: median estimate of $\theta_{12} = 0.3$, standard deviation (SD) of estimates, median standard error (SE), empirical power (%) to reject null at 5% significance, and empirical coverage (%) of 95% confidence interval

Shared variants	Total IVs	Median	SD	Median SE	Power (%)	Coverage (%)
Method 1—full set of interactions:						
0[a]	120	0.3080	0.0722	0.0718	98.8	95.2
1	119	0.3080	0.0723	0.0719	98.8	95.0
3	114	0.3090	0.0717	0.0716	98.9	95.3
5	105	0.3078	0.0716	0.0707	98.9	94.9
8	84	0.3073	0.0682	0.0687	99.3	95.2
10	65	0.3056	0.0670	0.0673	99.2	95.3
Method 2—reduced set of interactions:						
1	100	0.3073	0.0804	0.0794	96.7	94.9
3	66	0.3088	0.1003	0.0997	86.1	95.2
5	40	0.3056	0.1340	0.1334	63.2	95.7
8	16	0.3054	0.2520	0.2471	23.9	97.1
10	10	0.3057	0.3883	0.3891	8.7	99.3
Method 3—continuous gene scores:						
0	3	0.3000	0.1106	0.1091	77.5	95.8
1	3	0.3005	0.1111	0.1088	77.8	95.4
3	3	0.2998	0.1051	0.1048	81.0	95.6
5	3	0.3015	0.0997	0.0980	85.6	95.5
8	3	0.3003	0.0857	0.0858	93.0	95.8
10	3	0.2993	32.31	0.1711	42.7	99.2
Method 4—dichotomized gene scores:						
0	3	0.2972	0.1777	0.1722	41.8	96.0
1	3	0.3028	0.1757	0.1724	42.2	96.3
3	3	0.3002	0.1818	0.1773	39.8	96.4
5	3	0.3005	0.1948	0.1884	36.6	96.6
8	3	0.3007	0.2474	0.2340	25.7	97.2
10	3	0.2896	133.5	1.3578	0.7	100.0

[a]When there are no shared variants, methods 1 and 2 are equivalent.

variants) and Table 2 (varying the number of shared variants). All four approaches provided unbiased estimates of the interaction effect in all scenarios, with coverage for the 95% confidence interval close to the nominal 95% level. Power varied considerably between the methods. With no shared variants, method 1 (full set of interactions) and method 2 (reduced set of interactions) are equivalent and gave the most efficient estimates throughout. Method 3 (continuous gene scores) was less efficient, and method 4 (dichotomized gene scores) was the least efficient. With shared variants, method 1 was the most efficient throughout, and its efficiency was not strongly affected by the risk factors having genetic predictors in common. Between methods 2 and 3, method 2 was more efficient when most of the variants were non-shared, whereas method 3 was more efficient when most of the variants were shared. Again, method 4 was the least efficient in all scenarios. This suggests that the 2×2 approach may be underpowered in practice, and instead approaches using all genetic variants and their cross-terms should be considered.

We also varied the strength of the genetic variants due to potential concerns about weak instruments.[30] We considered scenarios in which the genetic variants explained 1% and 5% of variance in the risk factors. Although substantial weak instrument bias was observed for the main effects, no bias was observed for the interaction term, even when there were 100 IVs in the analysis and F-statistics and conditional F-statistics[31] for the product term were ~1

(Supplementary Tables A1 and A2, available as Supplementary data at *IJE* online). Similar findings were observed in a one-sample setting when varying the direction of confounder effects on the risk factor and outcome (results not shown). We also performed the simulation study centering the values of the risk factors to reduce the impact of collinearity. This changed the mean estimates of the main effects θ_1 and θ_2 and improved precision for the main effect estimates, but estimates and inferences for the interaction term θ_{12} were unchanged (Supplementary Table A3, available as Supplementary data at *IJE* online). These additional simulations suggest that factorial Mendelian randomization should only be used when the interaction is the main object of interest, and numerical estimates for the main effects from this model should be interpreted with caution.

Simulation study 2: interactions between interventions

Results from the simulation study for estimating interactions between the gene scores that act as proxies for the interventions are displayed in Table 3. Whereas the numerical values of estimates differed between the two approaches, a consistent finding was that power to detect an interaction was greater using continuous gene scores than using dichotomized gene scores. Varying the proportion of variance explained by the genetic variants had no discernable effect on the power to detect an interaction. This can be seen by comparing scenarios 1, 2 and 3, and scenarios 5 and 6. However, varying the minor allele frequency had a strong effect on power, with greater power when the minor allele frequency was close to 0.5. This can be seen by comparing scenarios 2, 4 and 5, and scenarios 3 and 6. This suggests that ensuring comparable size between subgroups is an important factor for efficient detection of interactions, and can be more important than ensuring that the strongest variant is used in the analysis.

Applied example: the effects of BMI and alcohol on systolic blood pressure

The lead variant (rs1229984) explained 0.24% of the variance in alcohol consumption, whereas the 10 variants

Table 3. Simulation study results for interaction between interventions: median estimate, standard deviation (SD) of estimates, median standard error (SE), and empirical power (%) to reject null at 5% significance. The minor allele frequencies and proportion of variance explained for variants that are proxies for interventions A and B are varied between scenarios

	Continuous gene scores				Dichotomized gene scores			
	Median	SD	Median SE	Power	Median	SD	Median SE	Power
Scenario 1: (A) common variants, 3%; (B) common variants, 3%								
θ_{12}=0.1	0.0583	0.0420	0.0417	29.3	0.0368	0.0423	0.0421	13.5
θ_{12}=0.3	0.0330	0.0080	0.0078	98.7	0.1102	0.0429	0.0423	73.5
θ_{12}=0.5	0.0224	0.0034	0.0032	100.0	0.1846	0.0428	0.0427	98.9
Scenario 2: (A) common variants, 5%; (B) common variants, 5%								
θ_{12}=0.1	0.0484	0.0343	0.0343	29.1	0.0372	0.0420	0.0422	13.5
θ_{12}=0.3	0.0304	0.0074	0.0072	98.8	0.1108	0.0424	0.0423	74.3
θ_{12}=0.5	0.0212	0.0033	0.0030	100.0	0.1851	0.0439	0.0427	99.0
Scenario 3: (A) common variants, 3%; (B) common variants, 7%								
θ_{12}=0.1	0.0498	0.0350	0.0352	29.2	0.0371	0.0422	0.0422	14.1
θ_{12}=0.3	0.0305	0.0075	0.0072	99.0	0.1106	0.0426	0.0423	74.2
θ_{12}=0.5	0.0213	0.0033	0.0030	100.0	0.1844	0.0430	0.0427	99.1
Scenario 4: (A) uncommon variants, 5%; (B) uncommon variants, 5%								
θ_{12}=0.1	0.0824	0.1152	0.1150	10.9	0.0168	0.0435	0.0430	7.0
θ_{12}=0.3	0.1082	0.0519	0.0500	58.8	0.0526	0.0434	0.0430	23.3
θ_{12}=0.5	0.0996	0.0300	0.0278	94.6	0.0879	0.0436	0.0430	53.0
Scenario 5: (A) common variants, 5%; (B) uncommon variants, 5%								
θ_{12}=0.1	0.0669	0.0699	0.0685	16.7	0.0246	0.0434	0.0425	9.1
θ_{12}=0.3	0.0618	0.0211	0.0204	85.5	0.0763	0.0433	0.0426	42.8
θ_{12}=0.5	0.0489	0.0109	0.0097	99.9	0.1279	0.0434	0.0428	84.1
Scenario 6: (A) common variants, 3%; (B) uncommon variants, 7%								
θ_{12}=0.1	0.0748	0.0756	0.0742	17.8	0.0259	0.0432	0.0426	9.7
θ_{12}=0.3	0.0649	0.0221	0.0215	85.4	0.0758	0.0430	0.0426	42.9
θ_{12}=0.5	0.0510	0.0113	0.0101	99.9	0.1271	0.0435	0.0428	83.9

explained 0.28% of the variance. Although the alcohol-decreasing allele of the rs1229984 variant is dominant, its frequency is only 2.5%. Dichotomizing participants based on this variant led to unequal groups in the population, whereas dichotomizing based on the 10 variant score led to equal groups (Table 4). However, the difference in mean alcohol levels between subgroups was reduced when using the 10 variant score, as most of the difference is due to the rs1229984 variant.

Estimates of the interaction between BMI and alcohol consumption are displayed in Table 5. For the dichotomized gene scores, efficiency is greater when the rs1229984 variant is used, suggesting the importance of dichotomizing the risk factor at a natural break in its distribution (if one exists) rather than ensuring that subgroups are equal in size. However, efficiency is strikingly improved using the full set of interactions, with the standard error decreasing over 10-fold using the 10 variants, and by a factor of 4 using the rs1229984 variant, compared with the 2×2 analysis. All estimates are compatible with the null, suggesting a lack of interaction in the effects of BMI and alcohol on SBP. There was no evidence of weak instrument bias, even though up to 857 IVs were used in the analyses and F-statistics were generally low (Supplementary Table A4, available as Supplementary data at *IJE* online).

Discussion

In this paper, we have provided a brief review of factorial Mendelian randomization, an approach that uses genetic variants as IVs to detect interactions. We have described two broad scenarios in which factorial Mendelian randomization has been implemented: to explore interactions between risk factors, and to explore interactions between interventions. Although most (perhaps even all) factorial Mendelian randomization analyses have been conducted using a 2×2 approach in which the sample is divided into four subgroups, we have shown that this approach is generally inefficient, particularly for exploring interactions between risk factors. This has been demonstrated in simulation studies, and in an applied example in which a 4- to 10-fold improvement in efficiency was

Table 5. Factorial Mendelian randomization results for applied example: estimates of interaction between BMI and alcohol consumption on systolic blood pressure; estimates are in mmHg units per 1 kg/m^2 change in BMI and 1 unit/day change in alcohol consumption

	Total IVs	Estimate	Standard error	*P*-value
10 variants for alcohol				
Method 1: full set of interactions	857	0.0023	0.0503	0.96
Method 2: continuous gene scores	3	0.0655	0.3402	0.85
Method 3: binary gene scores	3	0.1011	0.6411	0.87
rs1229984 variant for alcohol				
Method 1: full set of interactions	149	−0.0170	0.1136	0.88
Method 2: continuous gene scores	3	0.1917	0.3725	0.61
Method 3: binary gene scores	3	0.1499	0.4174	0.72

Table 4. Subgroups defined by genetic predictors of BMI and alcohol consumption: numbers (%) of participants and mean (standard deviation) of body mass index, alcohol consumption and systolic blood pressure in 2×2 subgroups when either 10 genetic variants or the rs1229984 variant used as IVs for alcohol consumption

		Mean (SD)		
	Participants (%)	BMI (kg/m^2)	Alcohol (units/day)	SBP (mmHg)
Overall	291, 781 (100.0)	27.1 (4.51)	2.54 (2.58)	140.0 (19.8)
10 variants for alcohol:				
Low BMI, low alcohol	73, 003 (25.0)	26.6 (4.25)	2.50 (2.52)	140.6 (20.6)
High BMI, low alcohol	72, 889 (25.0)	27.5 (4.65)	2.47 (2.50)	141.2 (20.6)
Low BMI, high alcohol	72, 888 (25.0)	26.7 (4.30)	2.61 (2.68)	140.8 (20.7)
High BMI, high alcohol	73, 001 (25.0)	27.6 (4.71)	2.59 (2.59)	141.3 (20.6)
rs1229984 variant for alcohol:				
Low BMI, low alcohol	6, 997 (2.4)	26.3 (4.10)	2.00 (2.04)	139.2 (20.2)
High BMI, low alcohol	6, 863 (2.4)	27.3 (4.50)	1.95 (1.99)	139.7 (20.2)
Low BMI, high alcohol	138, 894 (47.6)	26.7 (4.28)	2.59 (2.59)	140.8 (20.6)
High BMI, high alcohol	139, 027 (47.6)	27.6 (4.69)	2.56 (2.56)	141.3 (20.6)

observed by an analysis using the full set of interactions between the genetic variants as IVs.

Choice of variants

Our findings suggest that factorial Mendelian randomization analyses should be conducted using all available genetic variants that are valid instruments, i.e. that satisfy the multivariable IV assumptions. Analyses should not only include the genetic variants as main effects, but also all relevant two-way cross-terms. A similar conclusion was made in a different context by Bollen and Paxton.[22] If investigators want to perform a 2×2 analysis, this should be done to illustrate the method rather than being the main analysis for testing the presence of an interaction. For a 2×2 analysis, the primary consideration for choosing genetic variants should be to divide the population at a natural break in the distribution of the risk factor, in order to maximize the difference between the mean level of the risk factor in the two halves of the population. If there is no natural break in the distribution, then investigators should find a division that splits the population as far as possible into equal groups. This may entail selecting genetic variants that explain less variance in the risk factor, but have minor allele frequency closer to 50%. There can also be substantial benefit in including multiple variants in a single gene region in an analysis, even if these variants only explain a small additional proportion of variance in the risk factor.

Weak instrument bias and efficiency

Conventionally, it is discouraged to use large numbers of genetic variants that are not strongly associated with the risk factor in a Mendelian randomization analysis due to weak instrument bias.[32] Although we did not detect any bias from weak instruments on interaction terms in our simulations, we acknowledge that users of the method may be reluctant to use hundreds of cross-terms as IVs. We would therefore encourage the use of continuous gene score methods as sensitivity analyses. Such analyses estimate fewer parameters, so should be less susceptible to bias. However, this advice is precautionary; no evidence of weak instrument bias in interaction estimates was observed in our simulations.

Summarized data

Whereas multivariable Mendelian randomization can be performed using summarized data that are typically reported from genome-wide association studies by large consortia, this is not possible for factorial Mendelian randomization. If summarized association estimates are available on genetic associations with the product of the two risk factors, as well as associations with the risk factors individually, then the interaction effect can in principle be estimated by weighted linear regression of the beta-coefficients as in multivariable Mendelian randomization. However, if association estimates are only available for genetic variants, then the regression model is not identified asymptotically due to collinearity, and finite-sample estimates will be biased.[33] Association estimates for some cross-terms of genetic variants are additionally required. Hence, factorial Mendelian randomization can be performed using summarized data, but only if bespoke summarized data are available on associations of genetic variants and their cross-terms with the risk factors and their product.

Interpretation of the interaction effect

If genetic variants each satisfy the assumptions of an IV, then an interaction between risk factors has a causal interpretation. If the two risk factors are associated with the outcome then an interaction will exist on at least one of the additive or multiplicative scales.[6] However, there is no way of distinguishing a purely statistical interaction from a mechanistic or biological interaction based on observational data. We therefore advise caution in the interpretation of interaction findings, as a statistical interaction can arise due to non-linearity in the effect of a risk factor, or because of the scale on which the outcome is measured (for example, an interaction may occur on the original scale, but not on a log-transformed scale). When considering an interaction between interventions, researchers can investigate whether there is an interaction between the interventions on the risk factor(s) as well as on the outcome. This may help reveal where any biological interaction may take place.

Causal estimates from IV analysis have a clear interpretation in two cases: under the monotonicity assumption, and under the homogeneity assumption.[34] In a randomized controlled trial in which random allocation is taken as the IV and the treatment is the risk factor, monotonicity means that there are no individuals in the population (known as 'defiers') who would take the treatment only if they were randomly allocated to the control group, and not if they were allocated to the treatment group. Under monotonicity, all individuals are either 'always-takers' (they would always take the treatment whether assigned to or not), 'never-takers' (they would never take the treatment whether assigned to or not), or 'compliers' (they would take the treatment if and only if assigned to do so).[35] Under the monotonicity assumption, the IV estimate represents the complier average causal effect—the average

causal effect amongst compliers.[36] However, these definitions suppose that the IV and risk factor are binary. In Mendelian randomization, these variables are typically continuous, and so the straightforward interpretation of an IV estimate as a single complier average causal effect is lost—it instead represents a weighted average of complier average causal effects.[37] In contrast, the IV estimate under the homogeneity assumption represents the average causal effect. In its simplest form, the homogeneity assumption states that causal effects are identical in all individuals in the population. Weaker versions of this assumption have been proposed.

If there is a non-zero interaction between the risk factors, then the homogeneity assumption in the multivariable Mendelian randomization model is violated, and the IV estimate only has a clear interpretation under the monotonicity assumption. However, the homogeneity assumption in the factorial Mendelian randomization model may still hold, if there is homogeneity in the effects of the two risk factors and their product on the outcome. Hence under homogeneity, the interaction effect has an interpretation as an average causal effect.

A further potential complication arises if genetic associations with the risk factor or outcome vary over time. As genetic variants are assigned at conception for all individuals and tend to influence risk factor levels throughout the life-course, Mendelian randomization estimates are naturally interpreted as the impact of a life-long change in the trajectory of a risk factor.[38] Hence the natural interpretation of an interaction effect is that of a statistical interaction in the relationship between the outcome and the risk factors that relates to long-term changes in the risk factors. If genetic associations vary over time, then the interpretation of the causal estimate from Mendelian randomization is unclear. This is true for a conventional Mendelian randomization analysis as well as for a factorial Mendelian randomization analysis. One notable case to consider is if the risk factors have mutual effects on each other, as in the case of a feedback mechanism. In this situation, provided that the associations of the genetic variants with the risk factors remain linear (which would occur if all relationships between variables are linear), then this would mean that all genetic variants are associated with both risk factors. A factorial Mendelian randomization analysis would still hold for the causal interaction between the risk factors, as in the examples with shared genetic variants described earlier in the paper. Hence feedback between the risk factors does not necessarily lead to a non-zero interaction estimate. However, if the two variables of interest have a complex longitudinal relationship, and in particular if there are mutual dependencies that might vary over time, then extra caution should be taken in interpreting results from a Mendelian randomization investigation, especially numerical estimates of causal effects. This advice is also relevant if the effects of the risk factors on the outcome may vary over time (for example if there is a critical period when exposure to the risk factor influences the outcome). If the associations between variables became non-linear, then it may be worth considering using the control function approach, an extension to the two-stage least squares method that makes stronger assumptions, but can result in more efficient estimation.[39]

Comparison with previous work

Previous work investigating interactions using IVs has been limited. A formal framework for defining interaction effects in the context of clinical trials was proposed by Blackwell,[40] who used the language of principal stratification (compliance classes and monotonicity) to define local average interaction effects in a similar way to how local average causal effects (also called complier-averaged causal effects) are defined for single risk factors.[41] However, the principal stratification framework presupposes that risk factors are binary (or categorical) to assign compliance classes, whereas risk factors in Mendelian randomization are typically continuous. Additionally, the principal stratification framework presupposes a single binary IV, whereas Mendelian randomization investigations often use multiple genetic variants. There is therefore little practical advice in the literature on how to perform a factorial Mendelian randomization analysis.

Limitations

There are several limitations to this work. We rely on the assumption that all genetic variants included in our analyses are valid IVs. The IV assumptions may be violated by including genetic variants that are associated with the outcome independently of the risk factors. This violation would result in biased estimates, and could potentially lead to incorrect inferences on the presence of an interaction effect. Our recommendations rely on simulated data. Different choices for the parameters included in the simulation studies may have resulted in different conclusions. However, our findings were robust to different choices of parameters considered in this paper, they correspond to what we know about the theoretical properties of estimators, and similar conclusions were observed from the applied analysis. We have only considered interactions on an additive scale, although interactions could be considered on a multiplicative scale by log-transforming the outcome. Finally, we have not considered the impact of model misspecification on estimates. It would not be possible to

perform simulation studies corresponding to all possible ways that model misspecification could occur, meaning that our recommendations cannot be proven to be optimal in all settings. We believe that we have chosen parameters and scenarios that are relevant to modern Mendelian randomization analyses.

Conclusion

Overall, factorial Mendelian randomization is a promising technique for assessing interactions using genetic variants as IVs. Our findings suggest that current applications of factorial Mendelian randomization based on a 2×2 analysis could be improved by better selection of genetic variants, and by better choice of analysis method.

Supplementary data

Supplementary data are available at *IJE* online.

Funding

This work was supported by the UK Medical Research Council (MC_UU_00002/7) and the NIHR Cambridge Biomedical Research Centre. J.M.B.R. is supported by the British Heart Foundation (grant number FS/14/59/31282). S.B. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 204623/Z/16/Z). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Acknowledgement

This research has been conducted using the UK Biobank Resource under Application Number 7439.

Conflict of interest: None declared.

References

1. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22.
2. Burgess S, Thompson SG, *Mendelian Randomization: methods for Using Genetic Variants in Causal Estimation*. Boca Raton, FL: Chapman & Hall, 2015.
3. Lawlor D, Harbord R, Sterne J, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statist Med* 2008;**27**:1133–63.
4. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007;**16**:309–30.
5. Cox DR. Interaction. *Int Stat Rev* 1984;**52**:1–24.
6. VanderWeele TJ. *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York, NY: Oxford University Press; 2015.
7. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;**23**:R89–98.
8. Carter AR, Borges MC, Benn M *et al*. Combined association of body mass index and alcohol consumption with biomarkers for liver injury and incidence of liver disease: a Mendelian randomization study. *JAMA Netw Open* 2019;**2**:e190305.
9. Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in *NPC1L1*, *HMGCR*, or both: a 2×2 factorial Mendelian randomization study. *J Am Coll Cardiol* 2015;**65**:1552–61.
10. Ference BA, Robinson JG, Brook RD *et al*. Variation in *PCSK9* and *HMGCR* and risk of cardiovascular disease and diabetes. *N Engl J Med* 2016;**375**:2144–53.
11. Ference BA, Kastelein JJ, Ginsberg HN *et al*. Association of genetic variants related to CETP inhibitors and statins with lipoprotein levels and cardiovascular risk. *JAMA* 2017;**318**:947–56.
12. Stampfer MJ, Buring JE, Willett W, Rosner B, Eberlein K, Hennekens CH. The 2×2 factorial design: Its application to a randomized trial of aspirin and U.S. physicians. *Statist Med* 1985;**4**:111–16.
13. Hingorani A, Humphries S. Nature's randomised trials. *Lancet* 2005;**366**:1906–908.
14. Swanson SA, Tiemeier H, Ikram MA, Hernán MA. Nature as a trialist?: deconstructing the analogy between Mendelian randomization and randomized trials. *Epidemiology* 2017;**28**:653–59.
15. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015;**181**:251–60.
16. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single sample and two-sample summary data settings. *Int J Epidemiol* 2018.
17. Burgess S, Dudbridge F, Thompson SG. Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". *Am J Epidemiol* 2015;**181**:290–91.
18. McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. *J Am Med Assoc* 2003;**289**:2545–53.
19. Burgess S, Thompson DJ, Rees JM, Day FR, Perry JR, Ong KK. Dissecting causal pathways using Mendelian randomization with summarized genetic data: application to age at menarche and risk of breast cancer. *Genetics* 2017;**207**:481–7.
20. Dakin H, Gray A. Economic evaluation of factorial randomised controlled trials: challenges, methods and recommendations. *Statist Med* 2017;**36**:2814–30.
21. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data. Chapter 18: Estimating Average Treatment Effects*. Cambridge, MA: MIT Press; 2002.
22. Bollen KA, Paxton P, Two-stage least squares estimation of interaction effects. In: Schumacker RE, Marcoulides GA (eds). *Interaction and Nonlinear Effects in Structural Equation Modeling*. New Jersey: Lawrence Erlbaum Associates Publishers; 1998, pp.125–51.

23. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002;**360**:1903–13.
24. Wei GS, Coady SA, Goff DC *et al*. Blood pressure and the risk of developing diabetes in african americans and whites: ARIC, CARDIA, and the Framingham Heart Study. *Diabetes Care* 2011;**34**:873–79.
25. Droyvold WB, Midthjell K, Nilsen TI, Holmen J. Change in body mass index and its impact on blood pressure: a prospective population study. *Int J Obes* 2005;**29**:650–55.
26. Gelber RP, Gaziano JM, Manson JE, Buring JE, Sesso HD. A prospective study of body mass index and the risk of developing hypertension in men. *Am J Hypertens* 2007;**20**:370–77.
27. Roerecke M, Kaczorowski J, Tobe SW, Gmel G, Hasan OSM, Rehm J. The effect of a reduction in alcohol consumption on blood pressure: a systematic review and meta-analysis. *Lancet Public Health* 2017;**2**:e108–20.
28. Locke AE, Kahali B, Berndt SI *et al*. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;**518**: 197–206.
29. Lewis SJ, Zuccolo L, Smith GD *et al*. Fetal alcohol exposure and IQ at age 8: evidence from a population-based birth-cohort study. *PLOS One* 2012;**7**:e49407.
30. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Statist Med* 2011;**30**:1312–23.
31. Sanderson E, Windmeijer F. A weak instrument F-test in linear IV models with multiple endogenous variables. *J Econ* 2016; **190**:212–21.
32. Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* 2011;**40**:755–64.
33. Rees JMB. *Robust Methods in Mendelian Randomization. Chapter 5: Extending Mendelian Randomization to a Factorial Framework to Detect Interaction Effects*. Cambridge, UK: University of Cambridge; 2019.
34. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006;**17**:360–72.
35. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002;**58**:21–29.
36. Yau LHY, Little RJ. Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *J Am Stat Assoc* 2001;**96**:1232–44.
37. Angrist JD, Graddy K, Imbens GW. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Rev Econ Studies* 2000;**67**:499–527.
38. Labrecque JA, Swanson SA. Interpretation and potential biases of Mendelian randomization estimates with time-varying exposures. *Am J Epidemiol* 2019;**188**:231–38.
39. Guo Z, Small DS. Control function instrumental variable estimation of nonlinear causal effect models. *J Mach Learn Res* 2016; **17**:3448–82.
40. Blackwell M. Instrumental variable methods for conditional effects and causal interaction in voter mobilization experiments. *J Am Stat Assoc* 2017;**112**:590–99.
41. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica* 1994;**62**:467–75.

Appendix G

Appendix to paper 3

Appendix to the published paper contained in Appendix F.

Supplementary Material

In the Supplementary Material, we provide more detail on the two simulation studies and the applied example presented in the paper.

Simulation study 1: interactions between risk factors

The two risk factors X_1 and X_2 were generated for $i = 1, 2, \ldots, 10\,000$ participants from the following data-generating model:

$$X_{1i} = \sum_{j=1}^{J_1} \alpha_{1j} G_{1ji} + \sum_{j=1}^{J_c} \alpha_{1cj} G_{cji} + U_{1i} + \epsilon_{1i} \quad \text{and}$$
$$X_{2i} = \sum_{j=1}^{J_2} \alpha_{2j} G_{2ji} + \sum_{j=1}^{J_c} \alpha_{2cj} G_{cji} + U_{2i} + \epsilon_{2i}\,,$$

where $\boldsymbol{G_1}$ and $\boldsymbol{G_2}$ are the genetic variants associated with X_1 and X_2 respectively, and $\boldsymbol{G_c}$ are the set of shared variants that are associated with both X_1 and X_2 (bold font represents vectors). The genotypes (0, 1 or 2) were generated independently from binomial distributions $\text{Bin}(2, MAF_j)$, where MAF_j represents the minor allele frequency (MAF) of the j^{th} genetic variant, and was drawn from a uniform distribution $\text{Unif}(0.1, 0.5)$. $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_{1_c}}$ represent the effects of the genetic variants $\boldsymbol{G_1}$ and $\boldsymbol{G_c}$ on X_1, and $\boldsymbol{\alpha_2}$ and $\boldsymbol{\alpha_{2_c}}$ represent the effects of the genetic variants $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$ on X_2. The genetic associations were calculated so that $\boldsymbol{G_1}$ and $\boldsymbol{G_c}$, and $\boldsymbol{G_2}$ and $\boldsymbol{G_c}$, explained $\sigma_1^2 = \sigma_2^2 = 10\%$ of the variance in X_1 and X_2 respectively. To ensure that each genetic variant explained the same amount of variation in the risk factor, we rearranged:

$$\text{var}(G_{1j}) = \sigma_1^2 = 2 \times \alpha_{1j}^2 MAF_{1j}(1 - MAF_{1j}) \quad \text{and}$$
$$\text{var}(G_{2j}) = \sigma_2^2 = 2 \times \alpha_{2j}^2 MAF_{2j}(1 - MAF_{2j})\,,$$

to calculate the genetic associations:

$$\alpha_{1j} = \sqrt{\frac{\sigma_1^2/(J_1 + J_c)}{2 \times MAF_{1j}(1 - MAF_{1j})}}\,,$$
$$\alpha_{1cj} = \sqrt{\frac{\sigma_1^2/(J_1 + J_c)}{2 \times MAF_{cj}(1 - MAF_{cj})}}\,,$$
$$\alpha_{2j} = \sqrt{\frac{\sigma_2^2/(J_1 + J_c)}{2 \times MAF_{2j}(1 - MAF_{2j})}}\,,$$
$$\alpha_{2cj} = \sqrt{\frac{\sigma_2^2/(J_1 + J_c)}{2 \times MAF_{cj}(1 - MAF_{cj})}}\,.$$

U_1 and U_2 represent the set of confounding variables of the $X_1 - Y$ and $X_2 - Y$ associations. To ensure the confounders explained 25% of the variation in the risk factors, U_1 and U_2 were drawn independently from a normal distribution $\mathcal{N}(0, 0.25)$. To fix the variances of X_1 and X_2 to one, the error terms ϵ_1 and ϵ_2 were generated independently from a normal distribution

with mean zero, and variance:

$$\sigma^2_{\epsilon_1} = 1 - \sigma^2_1 - 0.25 \quad \text{and} \quad \sigma^2_{\epsilon_2} = 1 - \sigma^2_2 - 0.25\,.$$

The outcome Y was generated from:

$$Y_i = \theta_0 + \theta_1 X_{1i} + \theta_2 X_{2i} + \theta_{12} X_{12i} + 0.5U_{1i} + 0.5U_{2i} + \epsilon_{Yi}\,,$$

where θ_1 and θ_2 represent the main effects of X_1 and X_2 on Y, and θ_{12} represents the interaction effect of X_1 and X_2 on Y. X_{12} was generated by either: a) multiplying X_1 and X_2; or b) multiplying the mean centred values of the risk factors $(X_1 - \bar{X_1})$ and $(X_2 - \bar{X_2})$, where $\bar{X_1}$ and $\bar{X_2}$ are the mean values of X_1 and X_2. To ensure the risk factors and confounders explained less than a third of the variance in the outcome, the error term ϵ_Y was generated from a standard normal distribution $\mathcal{N}(0, 1)$.

Two-stage least squares regression models were fitted to either: a) the directly generated values of the risk factors (X_1, X_2, $X_{12} = X_1 \times X_2$); or b) the mean centred values of the risk factors ($X_1 - \bar{X_1}$, $X_2 - \bar{X_2}$, $X_{12} = (X_1 - \bar{X_1}) \times (X_2 - \bar{X_2})$). When the risk factors were mean centred, the model estimated the marginal effects θ_{1M} and θ_{2M} of X_1 and X_2 on Y, otherwise θ_1 and θ_2 were estimated. For example, when there were no shared variants $J_c = 0$, the marginal effects were approximately:

$$\begin{aligned}
\theta_{M1} &= \theta_1 + 0.3\theta_{12} + J_2\theta_{12}\left(\sqrt{\frac{0.1/J_2}{2 \times 0.3 \times 0.7}} \times 0.3 \times 2\right), \\
\theta_{M2} &= \theta_2 + 0.25\theta_{12} + J_1\theta_{12}\left(\sqrt{\frac{0.1/J_1}{2 \times 0.3 \times 0.7}} \times 0.3 \times 2\right).
\end{aligned} \tag{A1}$$

The genetic variants were either treated as individual IVs or as a single instrument in externally weighted gene scores GS_{X_1} and GS_{X_2} for X_1 and X_2. The external weights for the gene scores were based on an independent set of 10 000 individuals, and were produced from the same data generating model used for the main set of participants. The following four sets of genetic variants were used as IVs in separate two-stage least squares regression models:

- Method 1 – full set of interactions: the J_1, J_2 and J_c genetic variants used to generate X_1 and X_2, plus the unique interactions and quadratic terms of $(\boldsymbol{G_1} + \boldsymbol{G_c}) \times (\boldsymbol{G_2} + \boldsymbol{G_c})$.
- Method 2 – reduced set of interactions: the J_1, J_2 and J_c genetic variants used to generate X_1 and X_2, plus the interactions from the product $\boldsymbol{G_1} \times \boldsymbol{G_2}$.
- Method 3 – continuous gene scores: the two weighted gene scores GS_{X_1} and GS_{X_2}, and their product $GS_{X_1} \times GS_{X_2}$.
- Method 4 – dichotomized gene scores: the two dichotomized gene scores, and their product.

Method 1 represents the oracle model as it includes all of the variables used in the data generating model, whereas Methods 2 to 4 are misspecified and their performance should be compared to Method 1. In Method 2, we have included a subset of the cross-terms between the genetic variants to create a more realistic scenario where the full set of relevant IVs are not included in the analysis. Method 3 considers the impact of including all of the genetic variants into two separate weighted gene scores, and finally, Method 4 considers the impact of dichotomizing the weighted gene scores.

Data were generated 10 000 times with $\theta_0 = 0.2$, $\theta_1 = 0.3$, $\theta_2 = 0.2$, and $\theta_{12} = 0.1, 0.3$ and 0.5. Each risk factor was associated with $(J_1 + J_c) = (J_2 + J_c) = 10$ genetic variants, and the number of shared variants J_c was initially set to 0 to consider the scenario where none of the genetic variants were associated with risk factors (Table 1). The data were re-generated for $\sigma_1^2 = \sigma_2^2 = 5\%$ and 1%, for $J_c = 0$ (Supplementary Table A1) and $J_c = 5$ (Supplementary Table A2), and the analyses were re-performed on the directly generated values of the risk factors. Estimates of the F-statistic and conditional F-statistic for X_1, X_2 and X_{12} were recorded. The analyses were re-performed on the mean centred risk factors (Supplementary Table A3), and the number of shared variants was set to $J_c = 1, 3, 5, 8$ and 10 (Table 2). The following measurements were recorded for the estimates of θ_1, θ_2 and θ_{12}: median estimate; standard deviation of estimates; median standard error of estimates; empirical power at the 5% significance level; and empirical coverage of the 95% confidence interval. The conditional F-statistic (also known as the Sanderson–Windmeijer F-statistic [1]) represents the strength of the IVs for the risk factors in a joint model, and is the relevant measure of instrument strength for a multivariable Mendelian randomization analysis [2].

	F-stat	CF-stat	Median	SD	Median SE	Power	Coverage
Variants explain 10% of the variance in risk factors:							
Methods 1 & 2 [a] – full set of interactions							
$\theta_1 = 0.3$	10.3 (0.6)	2.1 (0.3)	0.3043	0.0918	0.0910	91.0	95.0
$\theta_2 = 0.2$	10.3 (0.6)	2.1 (0.3)	0.2034	0.0947	0.0945	57.9	95.5
$\theta_{12} = 0.3$	8.1 (0.6)	1.9 (0.2)	0.3080	0.0722	0.0718	98.8	95.2
Method 3 – continuous gene scores							
$\theta_1 = 0.3$	364.2 (23.4)	104.5 (25.6)	0.2998	0.1359	0.1332	61.9	95.6
$\theta_2 = 0.2$	364.5 (23.2)	103.9 (25.3)	0.2019	0.1405	0.1387	31.5	95.8
$\theta_{12} = 0.3$	273.7 (22.4)	97.8 (22.8)	0.3000	0.1106	0.1091	77.5	95.8
Method 4 - dichotomized gene scores							
$\theta_1 = 0.3$	224.2 (17.7)	41.9 (13.4)	0.3039	0.2145	0.2074	32.1	95.8
$\theta_2 = 0.2$	224.4 (17.7)	41.7 (13.3)	0.2047	0.2236	0.2164	15.2	96.2
$\theta_{12} = 0.3$	168.2 (16.3)	40.0 (12.4)	0.2972	0.1777	0.1722	41.8	96.0
Variants explain 5% of the variance in risk factors:							
Methods 1 & 2 [a] – full set of interactions							
$\theta_1 = 0.3$	5.4 (0.4)	1.5 (0.2)	0.3174	0.0931	0.0920	92.4	94.5
$\theta_2 = 0.2$	5.4 (0.4)	1.4 (0.2)	0.2166	0.0957	0.0959	62.0	94.8
$\theta_{12} = 0.3$	3.9 (0.4)	1.2 (0.2)	0.3087	0.0889	0.0888	92.8	95.0
Method 3 – continuous gene scores							
$\theta_1 = 0.3$	170.2 (15.5)	25.4 (11.7)	0.2988	0.2298	0.2121	29.9	96.9
$\theta_2 = 0.2$	170.1 (15.7)	25.2 (11.5)	0.1985	0.2421	0.2237	13.8	96.9
$\theta_{12} = 0.3$	109.4 (13.3)	23.8 (10.4)	0.3020	0.2458	0.2276	26.7	96.9
Method 4 - dichotomized gene scores							
$\theta_1 = 0.3$	107.3 (12.2)	10.7 (6.7)	0.2970	3.928	0.3367	12.6	98.9
$\theta_2 = 0.2$	106.9 (12.0)	10.6 (6.6)	0.1948	3.804	0.3551	5.4	98.7
$\theta_{12} = 0.3$	68.8 (10.2)	10.2 (6.1)	0.3033	4.065	0.3654	10.8	98.8
Variants explain 1% of the variance in risk factors:							
Methods 1 & 2 [a] – full set of interactions							
$\theta_1 = 0.3$	1.8 (0.2)	1.4 (0.2)	0.3681	0.0910	0.0901	97.7	88.4
$\theta_2 = 0.2$	1.8 (0.2)	1.4 (0.2)	0.2670	0.0930	0.0930	81.4	88.6
$\theta_{12} = 0.3$	1.4 (0.2)	1.0 (0.1)	0.3029	0.0971	0.0972	86.4	95.4
Method 3 – continuous gene scores							
$\theta_1 = 0.3$	29.5 (6.4)	1.9 (2.9)	0.2854	29.26	0.8411	2.8	99.9
$\theta_2 = 0.2$	29.4 (6.4)	1.9 (2.8)	0.1883	31.58	0.9203	1.0	99.9
$\theta_{12} = 0.3$	12.3 (4.1)	1.6 (2.1)	0.3185	52.32	1.537	0.7	100.0
Method 4 - dichotomized gene scores							
$\theta_1 = 0.3$	19.1 (5.1)	1.6 (2.8)	0.2992	123.8	1.063	1.9	99.9
$\theta_2 = 0.2$	19.0 (5.0)	1.5 (2.4)	0.1930	217.5	1.163	0.6	100.0
$\theta_{12} = 0.3$	8.1 (3.3)	1.2 (1.7)	0.3121	347.4	1.933	0.3	100.0

Supplementary Table A1: Simulation study results for interactions between risk factors varying the amount of variance in the risk factors explained by the genetic variants, with no shared variants and an interaction effect $\theta_{12} = 0.3$: mean F-statistic (F-stat), mean conditional F-statistic (CF-stat), median estimate, standard deviation (SD) of estimates, median standard error (SE), empirical power (%) to reject null at 5% significance, and empirical coverage (%) of 95% confidence interval.

[a] As there are no shared variants, methods 1 and 2 are equivalent.

	F-stat	CF-stat	Median	SD	Median SE	Power	Coverage
Variants explain 10% of the variance in risk factors:							
Method 1 – full set of interactions							
$\theta_1 = 0.3$	11.6 (0.7)	2.5 (0.4)	0.2981	0.0933	0.0927	89.1	95.0
$\theta_2 = 0.2$	11.6 (0.7)	2.5 (0.4)	0.1988	0.0955	0.0960	55.0	95.5
$\theta_{12} = 0.3$	13.4 (0.9)	2.2 (0.3)	0.3074	0.0707	0.0706	99.0	95.0
Method 2 – reduced set of interactions							
$\theta_1 = 0.3$	28.8 (1.8)	2.6 (0.4)	0.2970	0.1664	0.1649	44.2	95.8
$\theta_2 = 0.2$	28.8 (1.8)	2.6 (0.4)	0.1966	0.1719	0.1715	21.0	95.9
$\theta_{12} = 0.3$	32.6 (2.1)	2.3 (0.3)	0.3056	0.1337	0.1333	63.4	95.8
Method 3 – continuous gene scores							
$\theta_1 = 0.3$	366.4 (23.2)	131.8 (30.9)	0.2993	0.1272	0.1244	67.0	95.4
$\theta_2 = 0.2$	366.3 (23.4)	131.0 (30.7)	0.1992	0.1314	0.1293	35.1	95.4
$\theta_{12} = 0.3$	426.6 (29.1)	120.9 (26.9)	0.3008	0.1000	0.0978	84.8	95.4
Method 4 - dichotomized gene scores							
$\theta_1 = 0.3$	233.5 (18.1)	35.8 (12.4)	0.2984	0.2399	0.2302	25.9	96.4
$\theta_2 = 0.2$	233.5 (18.2)	35.6 (12.3)	0.2005	0.2482	0.2396	13.0	96.4
$\theta_{12} = 0.3$	284.1 (21.6)	33.8 (11.2)	0.3006	0.1950	0.1877	36.8	96.4
Variants explain 5% of the variance in risk factors:							
Method 1 – full set of interactions							
$\theta_1 = 0.3$	6.0 (0.5)	1.6 (0.2)	0.3052	0.0980	0.0983	87.7	95.2
$\theta_2 = 0.2$	6.0 (0.5)	1.5 (0.2)	0.2078	0.1018	0.1022	53.3	95.2
$\theta_{12} = 0.3$	6.1 (0.5)	1.3 (0.2)	0.3097	0.0925	0.0919	91.3	95.3
Method 2 – reduced set of interactions							
$\theta_1 = 0.3$	14.2 (1.2)	1.6 (0.3)	0.2982	0.1600	0.1588	48.4	96.3
$\theta_2 = 0.2$	14.2 (1.2)	1.6 (0.3)	0.1994	0.1665	0.1664	22.7	96.1
$\theta_{12} = 0.3$	13.9 (1.3)	1.4 (0.2)	0.3087	0.1621	0.1615	49.0	96.1
Method 3 – continuous gene scores							
$\theta_1 = 0.3$	171.8 (15.6)	32.9 (14.1)	0.3014	0.2078	0.1951	35.7	96.4
$\theta_2 = 0.2$	172.1 (15.4)	32.6 (13.9)	0.2041	0.2169	0.2043	16.9	96.5
$\theta_{12} = 0.3$	171.7 (17.6)	30.0 (12.0)	0.2981	0.2147	0.2010	32.6	96.5
Method 4 - dichotomized gene scores							
$\theta_1 = 0.3$	111.9 (12.5)	9.5 (6.4)	0.2933	0.8024	0.3732	10.2	99.1
$\theta_2 = 0.2$	112.2 (12.3)	9.4 (6.3)	0.1981	0.8127	0.3926	4.6	98.9
$\theta_{12} = 0.3$	117.6 (13.5)	8.9 (5.7)	0.3066	0.8619	0.3967	9.6	99.1
Variants explain 1% of the variance in risk factors:							
Method 1 – full set of interactions							
$\theta_1 = 0.3$	2.0 (0.2)	1.4 (0.2)	0.3504	0.0975	0.0971	94.4	92.0
$\theta_2 = 0.2$	2.0 (0.2)	1.3 (0.2)	0.2478	0.1003	0.1002	69.6	92.2
$\theta_{12} = 0.3$	1.6 (0.2)	1.0 (0.1)	0.3037	0.1051	0.1043	82.0	95.3
Method 2 – reduced set of interactions							
$\theta_1 = 0.3$	3.5 (0.6)	1.4 (0.2)	0.3225	0.1398	0.1395	63.8	95.6
$\theta_2 = 0.2$	3.5 (0.5)	1.4 (0.2)	0.2243	0.1459	0.1457	34.3	95.7
$\theta_{12} = 0.3$	2.6 (0.5)	1.1 (0.1)	0.3036	0.1771	0.1758	41.8	96.1
Method 3 – continuous gene scores							
$\theta_1 = 0.3$	31.0 (6.6)	2.5 (3.7)	0.2912	47.33	0.7448	3.6	99.9
$\theta_2 = 0.2$	30.9 (6.5)	2.3 (3.4)	0.1939	41.15	0.8014	1.1	99.9
$\theta_{12} = 0.3$	19.9 (5.4)	1.9 (2.4)	0.3030	72.69	1.315	0.6	99.9
Method 4 - dichotomized gene scores							
$\theta_1 = 0.3$	20.9 (5.3)	1.6 (2.9)	0.2967	65.97	1.108	1.5	99.9
$\theta_2 = 0.2$	20.8 (5.2)	1.5 (2.5)	0.1959	54.84	1.208	0.4	100.0
$\theta_{12} = 0.3$	14.1 (4.4)	1.2 (1.6)	0.3096	105.7	1.991	0.2	100.0

Supplementary Table A2: Simulation study results for interactions between risk factors varying the amount of variance in the risk factors explained by the genetic variants, with 5 shared variants and an interaction effect $\theta_{12} = 0.3$: mean F-statistic (F-stat), mean conditional F-statistic (CF-stat), median estimate, standard deviation (SD) of estimates, median standard error (SE), empirical power (%) to reject null at 5% significance, and coverage (%) of 95% confidence interval.

	Median	SD	Median SE	Power (%)	Coverage (%)
Methods 1 & 2 [a] – full set of interactions					
$\theta_1 = 0.3$	0.4311	0.0327	0.0320	100.0	-
$\theta_2 = 0.2$	0.3370	0.0328	0.0320	100.0	-
$\theta_{12} = 0.1$	0.1101	0.0721	0.0718	33.7	94.6
$\theta_1 = 0.3$	0.6679	0.0408	0.0320	100.0	-
$\theta_2 = 0.2$	0.5823	0.0413	0.0320	100.0	-
$\theta_{12} = 0.3$	0.3080	0.0722	0.0718	98.8	95.2
$\theta_1 = 0.3$	0.9044	0.0527	0.0320	100.0	-
$\theta_2 = 0.2$	0.8290	0.0528	0.0320	100.0	-
$\theta_{12} = 0.5$	0.5073	0.0715	0.0718	100.0	95.2
Method 3 – continuous gene scores					
$\theta_1 = 0.3$	0.4178	0.0348	0.0343	100.0	-
$\theta_2 = 0.2$	0.3234	0.0349	0.0343	100.0	-
$\theta_{12} = 0.1$	0.1010	0.1113	0.1091	15.4	95.5
$\theta_1 = 0.3$	0.6539	0.0424	0.0343	100.0	-
$\theta_2 = 0.2$	0.5691	0.0431	0.0343	100.0	-
$\theta_{12} = 0.3$	0.3000	0.1106	0.1091	77.5	95.8
$\theta_1 = 0.3$	0.8906	0.0539	0.0343	100.0	-
$\theta_2 = 0.2$	0.8165	0.0543	0.0343	100.0	-
$\theta_{12} = 0.5$	0.4995	0.1107	0.1092	98.7	95.6
Method 4 – dichotomized gene scores					
$\theta_1 = 0.3$	0.4173	0.0438	0.0435	100.0	-
$\theta_2 = 0.2$	0.3236	0.0438	0.0434	100.0	-
$\theta_{12} = 0.1$	0.1022	0.1786	0.1720	8.0	95.9
$\theta_1 = 0.3$	0.6538	0.0496	0.0435	100.0	-
$\theta_2 = 0.2$	0.5687	0.0506	0.0435	100.0	-
$\theta_{12} = 0.3$	0.2972	0.1777	0.1722	41.8	96.0
$\theta_1 = 0.3$	0.8913	0.0597	0.0435	100.0	-
$\theta_2 = 0.2$	0.8165	0.0603	0.0435	100.0	-
$\theta_{12} = 0.5$	0.5002	0.1776	0.1718	80.7	96.1

Supplementary Table A3: Simulation study results for interactions between risk factors with no shared variants after centering the risk factors: median estimate, standard deviation (SD) of estimates, median standard error (SE), empirical power (%) to reject null at 5% significance, and empirical coverage (%) of 95% confidence interval. Note that centering changes the estimands for the main effect terms, not only the estimates – hence coverage is only displayed for the interaction term.

[a] As there are no shared variants, methods 1 and 2 are equivalent.

Simulation study 2: interactions between interventions

Using the same notation defined in the first simulation study, the risk factor X was generated for $i = 1, 2, \ldots, 10\,000$ participants from the following data generating model:

$$X_i = 0.3 + \sum_{j=1}^{J_A} \alpha_{Aj} G_{Aji} + \sum_{j=1}^{J_B} \alpha_{Bj} G_{Bji} + \alpha_{AB} \sum_{j=1}^{J_A \times J_B} G_{ABji} + U_i + \epsilon_{Xi}\,.$$

We assume that the two gene regions are distinct, and the genetic variants $\boldsymbol{G_A}$ and $\boldsymbol{G_B}$ are not in linkage disequilibrium. The genotypes were generated independently from binomial distributions $\text{Bin}(2, MAF_j)$, where MAF_j represents the MAF for the j^{th} genetic variant. MAF_j was drawn from a uniform distribution $\mathcal{U}(MAF_L, MAF_U)$, where the value of MAF_L and MAF_U were either taken as 0.4 and 0.5 (common variants), or 0.1 and 0.2 (uncommon variants). We assumed that the interaction effect α_{AB} was constant across the $J_A \times J_B$ product terms for simplicity.

The approximate proportion of variance explained in X by $\boldsymbol{G_A}$ (σ^2_A) and $\boldsymbol{G_B}$ (σ^2_B) varied between scenarios. As before, the genetic associations $\boldsymbol{\alpha_A}$ and $\boldsymbol{\alpha_B}$ were calculated by rearranging the formula for the variance of the genetic variants to ensure the amount of variance explained by each variant was the same:

$$\alpha_{Aj} = \sqrt{\frac{\sigma^2_A / J_A}{2 \times MAF_{Aj}(1 - MAF_{Aj})}} \quad \text{and}$$

$$\alpha_{Bj} = \sqrt{\frac{\sigma^2_B / J_B}{2 \times MAF_{Bj}(1 - MAF_{Bj})}}\,.$$

The confounders U were drawn from $\mathcal{N}(0, 0.25)$, and the error term ϵ_X was generated from $\mathcal{N}(0, 0.65)$. The outcome Y was generated from:

$$Y_i = \theta_0 + \theta_1 X_i + U_i + \epsilon_{Yi}\,,$$

where θ_1 represents the causal effect of X on Y, and the error term ϵ_Y was generated from a standard normal distribution $\mathcal{N}(0, 1)$. The data was generated 10 000 times under the following scenarios:

- Scenario 1: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.4, 0.5)$, $\sigma^2_A = 3\%$ and $\sigma^2_B = 3\%$
- Scenario 2: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.4, 0.5)$, $\sigma^2_A = 5\%$ and $\sigma^2_B = 5\%$
- Scenario 3: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.4, 0.5)$, $\sigma^2_A = 3\%$ and $\sigma^2_B = 7\%$
- Scenario 4: $MAF_A \sim \mathcal{U}(0.1, 0.2)$, $MAF_B \sim \mathcal{U}(0.1, 0.2)$, $\sigma^2_A = 5\%$ and $\sigma^2_B = 5\%$
- Scenario 5: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.1, 0.2)$, $\sigma^2_A = 5\%$ and $\sigma^2_B = 5\%$
- Scenario 6: $MAF_A \sim \mathcal{U}(0.4, 0.5)$, $MAF_B \sim \mathcal{U}(0.1, 0.2)$, $\sigma^2_A = 3\%$ and $\sigma^2_B = 7\%$

with $J_A = J_B = 3$, $\theta_0 = 0.2$, $\theta_1 = 0.1$, and $\alpha_{AB} = 0.1, 0.3$ and 0.5. The above scenarios were selected to consider the impact of varying the MAF and the amount of variance in the risk factor explained by the genetic variants had on the performance of the method.

For each scenario, optimal weighted gene scores GS_A and GS_B were generated for each gene region, where the external weights were produced from an independent set of 10 000

individuals from the same data-generating model used for the main set of participants. The two gene scores were dichotomized at their median values to create two binary variables. The outcome was then regressed against: a) the two continuous gene scores and their product; and b) the dichotomized gene scores and their product. The following measurements were recorded for the estimate of the interaction effect between the gene scores on the outcome: median estimate; standard deviation of estimates; median standard error; and empirical power at the 5% significance level.

Applied example: the effects of BMI and alcohol on systolic blood pressure

UK Biobank is a prospective, population-based cohort consisting of approximately 500,000 participants aged between 40 and 69 years at baseline living in the UK. Extensive baseline characteristics were collected at recruitment, including lifestyle factors, sociodemographic information, and physical attributes. For the analysis, we considered 367,643 unrelated participants of European descent who passed data quality control measures and had genetic data [3].

Body mass index (BMI, kg/m^2) and systolic blood pressure (SBP, mmHg) were measured at baseline when participants attended the assessment centre. Information on baseline alcohol consumption was obtained from a touchscreen questionnaire which included questions on alcohol drinking status, frequency of alcohol consumption, and beverage type. The responses to the amount of alcohol drank and beverage type were used to create a continuous variable that represented alcohol consumption in units per day. To adjust for blood pressure medication, 15 mmHg was added to SBP for individuals who reported to be on blood pressure lowering medication [4]. Individuals were dropped from the analysis if they had missing data on BMI, SBP, alcohol consumption, or relevant genetic variants. The final sample size was 291,781.

We used the 77 genome-wide significant variants from a meta-analysis by the Genetic Investigation of ANthropometric Traits (GIANT) consortium in participants of European ancestry to act as IVs for BMI [5]. For alcohol, we identified 10 genetic variants in the *ADH1B* gene region that have been shown to be associated with alcohol consumption [6]. The genetic variants used as IVs for BMI and alcohol consumption were cross-referenced to check for any overlap. BMI was regressed separately against each of the 10 alcohol variants, and alcohol consumption was regressed against each of the 77 BMI variants. All models were adjusted for gender, age, and the first ten genomic principal components.

Internally-weighted gene scores were created for BMI based on the 77 genetic variants (GS_{BMI}), and for alcohol consumption based on the 10 genetic variants (GS_{AC}), and these gene scores were dichotomized at their median values to create two binary variables. A separate binary variable was generated using the rs1229984 variant only, where participants were either considered to have: a) a low alcohol consumption if they were homozygous or heterozygous for the alcohol-decreasing allele; or b) a high alcohol consumption if they were homozygous for the alcohol-increasing allele (as in the paper by Carter *et al.* [7]). Using these binary variables, the following groups of participants were created:

- Low BMI, low alcohol consumption: $GS_{BMI} \leq med(GS_{BMI})$ and $GS_{AC} \leq med(GS_{AC})$ or was homozygous or heterozygous for the alcohol decreasing allele for the rs1229984 variant,
- High BMI, low alcohol consumption: $GS_{BMI} > med(GS_{BMI})$ and $GS_{AC} \leq med(GS_{AC})$ or was homozygous or heterozygous for the alcohol decreasing allele for the rs1229984 variant,
- Low BMI, high alcohol consumption: $GS_{BMI} \leq med(GS_{BMI})$ and $GS_{AC} > med(GS_{AC})$ or was homozygous for the alcohol increasing allele for the rs1229984 variant, and
- High BMI, high alcohol consumption: $GS_{BMI} \leq med(GS_{BMI})$ and $GS_{AC} > med(GS_{AC})$ or was homozygous for the alcohol increasing allele for the rs1229984 variant.

The above criteria created four groups of participants based on the dichotomized gene scores for BMI and alcohol consumption, and another four groups based on the dichotomized gene

score for BMI and the rs1229984 variant. The numbers of participants, and the mean and standard deviation of BMI, alcohol consumption, and SBP were recorded for each group.

Two-stage least squares regression models of SBP were fitted to BMI, alcohol consumption, and the product of BMI and alcohol consumption. The following sets of IVs were considered:

- Method 1: the 77 variants for BMI and 10 variants for alcohol consumption, plus 770 cross-terms between the two sets of variants.
- Method 2: the continuous gene scores GS_{BMI} and GS_{AC}, plus their product $GS_{BMI} \times GS_{AC}$.
- Method 3: the dichotomized gene scores of GS_{BMI} and GS_{AC}, plus their product.

The models were refitted excluding all of the variants for alcohol consumption apart from the lead rs1229984 variant. All models were adjusted for gender, age, and the first ten genomic principal components. For each model, the estimate and standard error of the interaction term was recorded with its p-value. In total, six two-stage least squares regression models were fitted to the dataset, and all of the models were adjusted for age, gender and the first 10 genomic principal components. The F-statistic and the Sanderson–Windmeijer conditional F-statistic were estimated for each set of IVs with respect to BMI, alcohol consumption, and the product of BMI and alcohol consumption (Supplementary Table A4).

	Method 1		Method 2		Method 3	
	F-stat	CF-stat	F-stat	CF-stat	F-stat	CF-stat
10 variants for alcohol:						
BMI	6.8	1.3	1662.8	21.1	1054.1	7.0
Alcohol consumption	2.4	1.1	268.0	20.9	55.6	6.9
Product term	2.4	1.1	298.6	21.0	73.2	6.9
rs1229984 for alcohol:						
BMI	32.8	1.3	1654.9	17.2	1066.8	13.5
Alcohol consumption	7.7	1.2	245.1	17.1	241.6	13.4
Product term	7.9	1.2	267.7	17.1	266.5	13.4

Supplementary Table A4: F-statistics (F-stat) and conditional F-statistics (CF-stat) for applied example.

Supplementary References

[1] Sanderson E, Windmeijer F. A weak instrument F-test in linear IV models with multiple endogenous variables. Journal of Econometrics. 2016;190(2):212–221.

[2] Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single sample and two-sample summary data settings. International Journal of Epidemiology. 2018;.

[3] Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell. 2016;167(5):1415–1429.

[4] Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nature Genetics. 2017;49(3):403–415.

[5] Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197–206.

[6] Lewis SJ, Zuccolo L, Smith GD, Macleod J, Rodriguez S, Draper ES, et al. Fetal alcohol exposure and IQ at age 8: evidence from a population-based birth-cohort study. PLOS One. 2012;7(11):e49407.

[7] Carter AR, Borges MC, Benn M, Tybjærg-Hansen A, Davey Smith G, Nordestgaard BG, et al. Combined association of body mass index and alcohol consumption with biomarkers for liver injury and incidence of liver disease: a Mendelian randomization study. JAMA Network Open. 2019;2(3):e190305.

Appendix H

Chapter 3 supplementary material

H.1 Non-convergent robust regression models

The number of robust regression models that did not converge in the simulation study outlined in Section 3.5 are contained in Table H.1. Robust regression was applied to the IVW and MR-Egger models, and was also considered in conjunction with penalized weights.

Table H.1 Number of the 10 000 simulations that failed to report a standard error using robust regression (without and with penalized weights) with the inverse-variance weighted (IVW) and MR-Egger methods, for Scenarios 1-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instrumental variables.

	IVW								MR-Egger							
	Robust				Robust, penalized				Robust				Robust, penalized			
No. invalid:	0	1	3	6	0	1	3	6	0	1	3	6	0	1	3	6
Null causal effect: $\theta = 0$																
Scenario 1	0	-	-	-	0	-	-	-	16	-	-	-	16	-	-	-
Scenario 2	-	1	2	5	-	3	9	120	-	24	72	78	-	45	98	258
Scenario 3	-	2	1	4	-	3	10	51	-	32	69	32	-	30	70	139
Scenario 4	-	3	84	11	-	5	6	22	-	144	100	5	-	124	76	9
Positive causal effect: $\theta = 0.3$																
Scenario 1	4	-	-	-	3	-	-	-	13	-	-	-	13	-	-	-
Scenario 2	-	0	0	1	-	1	3	54	-	20	55	47	-	24	71	211
Scenario 3	-	0	0	3	-	3	2	22	-	24	72	19	-	37	62	73
Scenario 4	-	2	30	4	-	0	9	9	-	151	91	10	-	122	81	15

Abbreviations: IVW, inverse-variance weighted; No., number.

H.2 MR-Egger models

Table H.2 contains the results from the simulation study (Section 3.5) when the MR-Egger model was applied to the simulated data with: 1) robust regression (Rr); 2) penalized weights (PW); and 3) robust regression and penalized weights (Rr and PW).

Table H.2 Mean (standard error) estimates and power from the MR-Egger method with: robust regression (Rr); penalized weights (PW); and robust regression and penalized weights (Rr and PW) for Scenarios 1-4 with a null ($\theta = 0$) or positive ($\theta = 0.3$) causal effect by the number of invalid instrumental variables.

	No invalid IVs		1 invalid IV		3 invalid IVs		6 invalid IVs	
	Mean (mean SE)	Power, %	Mean (mean SE)	Power, %	Mean (mean SE)	Power, %	Mean (mean SE)	Power, %
Null causal effect: $\theta = 0$								
Scenario 1. No pleiotropy, InSIDE satisfied								
Rr	0.000 (0.231)	8.2	-	-	-	-	-	-
PW	-0.001 (0.216)	4.3	-	-	-	-	-	-
Rr and PW	0.000 (0.230)	8.3	-	-	-	-	-	-
Scenario 2. Balanced pleiotropy, InSIDE satisfied								
Rr	-	-	-0.006 (0.245)	9.7	-0.002 (0.375)	9.6	-0.007 (0.671)	10.8
PW	-	-	-0.006 (0.208)	9.9	-0.003 (0.231)	16.9	-0.009 (0.274)	31.5
Rr and PW	-	-	-0.007 (0.254)	9.2	-0.001 (0.333)	10.7	-0.008 (0.505)	20.3
Scenario 3. Directional pleiotropy, InSIDE satisfied								
Rr	-	-	-0.003 (0.246)	9.9	0.001 (0.376)	9.5	-0.004 (0.564)	13.8
PW	-	-	-0.004 (0.208)	10.1	0.001 (0.249)	18.7	-0.009 (0.343)	37.9
Rr and PW	-	-	-0.004 (0.256)	9.4	0.001 (0.309)	12.0	-0.005 (0.419)	32.4
Scenario 4. Directional pleiotropy, InSIDE violated								
Rr	-	-	0.171 (0.291)	18.2	0.493 (0.234)	65.1	0.649 (0.158)	95.6
PW	-	-	0.241 (0.196)	33.0	0.527 (0.178)	81.2	0.651 (0.159)	97.7
Rr and PW	-	-	0.173 (0.272)	18.5	0.490 (0.215)	68.2	0.652 (0.148)	96.8
Positive causal effect: $\theta = 0.3$								
Scenario 1. No pleiotropy, InSIDE satisfied								
Rr	0.144 (0.273)	13.1	-	-	-	-	-	-
PW	0.143 (0.258)	7.9	-	-	-	-	-	-
Rr and PW	0.144 (0.271)	13.3	-	-	-	-	-	-
Scenario 2. Balanced pleiotropy, InSIDE satisfied								
Rr	-	-	0.140 (0.295)	13.1	0.139 (0.430)	11.6	0.124 (0.665)	11.9
PW	-	-	0.139 (0.255)	13.5	0.140 (0.282)	19.4	0.130 (0.331)	30.2
Rr and PW	-	-	0.140 (0.297)	12.9	0.140 (0.363)	14.6	0.133 (0.508)	22.4
Scenario 3. Directional pleiotropy, InSIDE satisfied								
Rr	-	-	0.141 (0.295)	13.2	0.135 (0.433)	11.8	0.136 (0.563)	14.7
PW	-	-	0.140 (0.252)	13.2	0.137 (0.302)	19.9	0.137 (0.392)	30.7
Rr and PW	-	-	0.140 (0.292)	13.3	0.135 (0.352)	15.9	0.138 (0.433)	31.0
Scenario 4. Directional pleiotropy, InSIDE violated								
Rr	-	-	0.338 (0.340)	25.5	0.719 (0.274)	75.6	0.893 (0.190)	97.7
PW	-	-	0.418 (0.233)	48.9	0.754 (0.210)	91.2	0.895 (0.188)	99.4
Rr and PW	-	-	0.340 (0.319)	27.1	0.716 (0.249)	78.6	0.897 (0.179)	98.2

Abbreviations: IV, instrumental variable; SE, standard error; InSIDE, instrument strength independent of direct effect; Rr, robust regression; PW, penalized weights.

H.3 One-sample Mendelian randomization

Results from the simulation study (Section 3.5) when the data was generated from one sample are contained in Table H.3 (null causal effect $\theta = 0$) and Table H.4 (positive causal effect $\theta = 0.3$). Estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (Rr); 3) penalized weights (PW); 4) robust regression and penalized weights (Rr and PW); and 5) the three sets of variants selected by the least trimmed squared (LTS) estimator, and the Lasso selection (LS) method with the heterogeneity stopping rule for Scenarios 1-4 are displayed in the Tables H.3 and H.4.

Table H.3 Mean (standard error) and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (Rr); 3) penalized weights (PW); 4) robust regression and penalized weights (Rr and PW); and 5) the three sets of variants selected by the least trimmed squared (LTS) estimator for Scenarios 1-4 with a null causal effect ($\theta = 0$) by the number of invalid instrumental variables for one-sample Mendelian randomization. Results from the Lasso selection (LS) method with the heterogeneity stopping rule are also provided.

	No invalid IVs		1 invalid IV		3 invalid IVs		6 invalid IVs	
	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %
Null causal effect: $\boldsymbol{\theta = 0}$								
Scenario 1. No pleiotropy, InSIDE satisfied								
IVW	0.021 (0.061)	5.6	-	-	-	-	-	-
Rr	0.021 (0.065)	6.8	-	-	-	-	-	-
PW	0.019 (0.060)	6.0	-	-	-	-	-	-
Rr and PW	0.020 (0.063)	7.2	-	-	-	-	-	-
LTS[a]								
h	0.014 (0.078)	18.9	-	-	-	-	-	-
$w_{LTS,2}$	0.019 (0.061)	7.5	-	-	-	-	-	-
Auto	0.021 (0.060)	5.9	-	-	-	-	-	-
LS	0.021 (0.060)	6.2	-	-	-	-	-	-
Scenario 2. Balanced pleiotropy, InSIDE satisfied								
IVW	-	-	0.020 (0.088)	6.4	0.020 (0.132)	7.1	0.024 (0.180)	7.3
Rr	-	-	0.021 (0.068)	7.8	0.020 (0.096)	6.8	0.023 (0.195)	5.9
PW	-	-	0.018 (0.062)	7.1	0.015 (0.066)	9.7	0.008 (0.075)	19.7
Rr and PW	-	-	0.019 (0.070)	6.8	0.017 (0.092)	6.3	0.010 (0.156)	7.7
LTS[a]								
h	-	-	0.014 (0.078)	18.5	0.013 (0.079)	17.8	0.012 (0.08)	15.6
$w_{LTS,2}$	-	-	0.019 (0.063)	7.4	0.019 (0.075)	9.0	0.022 (0.142)	15.2
Auto	-	-	0.020 (0.063)	6.9	0.020 (0.071)	9.3	0.019 (0.091)	17.8
LS	-	-	0.020 (0.063)	7.1	0.020 (0.070)	9.2	0.019 (0.088)	16.9
Scenario 3. Directional pleiotropy, InSIDE satisfied								
IVW	-	-	0.086 (0.088)	9.3	0.216 (0.123)	36.2	0.409 (0.150)	92.5
Rr	-	-	0.032 (0.067)	8.8	0.088 (0.109)	11.2	0.357 (0.222)	44.9
PW	-	-	0.025 (0.062)	7.0	0.046 (0.067)	13.6	0.132 (0.081)	40.4
Rr and PW	-	-	0.025 (0.070)	7.5	0.040 (0.088)	10.8	0.103 (0.125)	21.5
LTS[a]								
h	-	-	0.015 (0.078)	18.6	0.018 (0.079)	16.8	0.035 (0.08)	14.6
$w_{LTS,2}$	-	-	0.024 (0.063)	6.9	0.059 (0.076)	13.0	0.306 (0.129)	71.1
Auto	-	-	0.027 (0.063)	6.9	0.049 (0.072)	12.6	0.140 (0.095)	36.0
LS	-	-	0.027 (0.063)	7.2	0.049 (0.071)	12.6	0.173 (0.096)	42.8
Scenario 4. Directional pleiotropy, InSIDE violated								
IVW	-	-	0.096 (0.068)	27.5	0.202 (0.072)	86.6	0.303 (0.067)	100.0
Rr	-	-	0.053 (0.081)	10.1	0.163 (0.119)	38.4	0.302 (0.072)	98.4
PW	-	-	0.040 (0.061)	14.2	0.095 (0.062)	41.4	0.237 (0.061)	89.6
Rr and PW	-	-	0.038 (0.069)	11.4	0.089 (0.079)	30.7	0.236 (0.071)	83.9
LTS[a]								
h	-	-	0.021 (0.078)	20.1	0.039 (0.076)	24.6	0.169 (0.067)	53.1
$w_{LTS,2}$	-	-	0.038 (0.062)	13.7	0.113 (0.067)	47.9	0.282 (0.066)	95.8
Auto	-	-	0.049 (0.062)	17.4	0.132 (0.066)	58.9	0.285 (0.065)	95.6
LS	-	-	0.048 (0.062)	17.2	0.138 (0.066)	58.8	0.300 (0.064)	99.1

Abbreviations: IV, instrumental variable; SE, standard error; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted; Rr, robust regression; PW, penalized weights; LTS, least trimmed squares; LS, Lasso selection; Auto, automated.

[a]The following three sets of genetic variants were selected from the LTS estimator and included in the IVW model: 1) the h=8 variants used to estimate the initial LTS estimate $\hat{\theta}_{LTS,h}$; 2) the variants with a weight of 1 in $w_{LTS,2}$; and 3) the variants selected from the automated approach based on the heterogeneity stopping criteria.

Table H.4 Mean (standard error) and power (%) of the estimates from the IVW model with: 1) the J genetic variants (IVW); 2) robust regression (Rr); 3) penalized weights (PW); 4) robust regression and penalized weights (Rr and PW); and 5) the three sets of variants selected by the least trimmed squared (LTS) estimator for Scenarios 1-4 with a positive causal effect ($\theta = 0.3$) by the number of invalid instrumental variables for one-sample Mendelian randomization. Results from the Lasso selection (LS) method with the heterogeneity stopping rule are also provided.

	No invalid IVs		1 invalid IV		3 invalid IVs		6 invalid IVs	
	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %	Mean (mean SE)	Pow., %
Positive causal effect: $\boldsymbol{\theta = 0.3}$								
Scenario 1. No pleiotropy, InSIDE satisfied								
IVW	0.321 (0.068)	99.7	-	-	-	-	-	-
Rr	0.321 (0.073)	97.7	-	-	-	-	-	-
PW	0.321 (0.068)	99.7	-	-	-	-	-	-
Rr and PW	0.321 (0.072)	97.9	-	-	-	-	-	-
LTS[a]								
h	0.312 (0.092)	84.8	-	-	-	-	-	-
$w_{LTS,2}$	0.319 (0.069)	99.0	-	-	-	-	-	-
Auto	0.321 (0.068)	99.7	-	-	-	-	-	-
LS	0.321 (0.068)	99.7	-	-	-	-	-	-
Scenario 2. Balanced pleiotropy, InSIDE satisfied								
IVW	-	-	0.322 (0.090)	91.2	0.322 (0.133)	66.2	0.323 (0.180)	43.9
Rr	-	-	0.322 (0.073)	97.2	0.321 (0.098)	86.1	0.323 (0.193)	43.6
PW	-	-	0.320 (0.070)	99.1	0.316 (0.076)	96.4	0.308 (0.087)	85.4
Rr and PW	-	-	0.321 (0.078)	96.3	0.317 (0.096)	88.4	0.310 (0.138)	67.3
LTS[a]								
h	-	-	0.315 (0.092)	85.7	0.314 (0.093)	86.0	0.312 (0.093)	87.8
$w_{LTS,2}$	-	-	0.320 (0.071)	98.9	0.320 (0.081)	94.4	0.321 (0.142)	60.1
Auto	-	-	0.322 (0.071)	99.2	0.320 (0.079)	95.3	0.320 (0.103)	78.8
LS	-	-	0.322 (0.071)	99.2	0.320 (0.079)	95.6	0.320 (0.099)	81.4
Scenario 3. Directional pleiotropy, InSIDE satisfied								
IVW	-	-	0.386 (0.088)	99.7	0.517 (0.124)	100.0	0.710 (0.150)	100.0
Rr	-	-	0.332 (0.073)	97.9	0.390 (0.111)	93.3	0.655 (0.226)	86.8
PW	-	-	0.330 (0.070)	99.7	0.362 (0.076)	99.5	0.465 (0.093)	99.5
Rr and PW	-	-	0.328 (0.078)	96.0	0.351 (0.093)	92.6	0.434 (0.125)	89.1
LTS[a]								
h	-	-	0.316 (0.092)	86.2	0.320 (0.093)	86.9	0.336 (0.093)	91.1
$w_{LTS,2}$	-	-	0.324 (0.071)	99.1	0.361 (0.082)	99.1	0.602 (0.130)	99.3
Auto	-	-	0.331 (0.071)	99.6	0.365 (0.080)	99.4	0.499 (0.108)	98.2
LS	-	-	0.331 (0.071)	99.6	0.363 (0.079)	99.4	0.508 (0.108)	99.2
Scenario 4. Directional pleiotropy, InSIDE violated								
IVW	-	-	0.396 (0.070)	100.0	0.502 (0.073)	100.0	0.601 (0.067)	100.0
Rr	-	-	0.352 (0.088)	95.5	0.463 (0.125)	89.7	0.601 (0.073)	99.7
PW	-	-	0.349 (0.068)	99.7	0.424 (0.068)	99.7	0.561 (0.064)	99.9
Rr and PW	-	-	0.343 (0.078)	97.0	0.414 (0.090)	95.7	0.562 (0.071)	98.8
LTS[a]								
h	-	-	0.319 (0.092)	86.3	0.339 (0.090)	87.4	0.468 (0.079)	94.5
$w_{LTS,2}$	-	-	0.336 (0.070)	99.3	0.415 (0.071)	99.1	0.580 (0.067)	99.6
Auto	-	-	0.358 (0.068)	99.7	0.462 (0.070)	99.5	0.595 (0.066)	99.7
LS	-	-	0.357 (0.068)	99.7	0.463 (0.070)	99.8	0.600 (0.066)	100.0

Abbreviations: IV, instrumental variable; SE, standard error; Pow., power; InSIDE, instrument strength independent of direct effect; IVW, inverse variance weighted; Rr, robust regression; PW, penalized weights; LTS, least trimmed squares; LS, Lasso selection; Auto, automated.

[a]The following three sets of genetic variants were selected from the LTS estimator and included in the IVW model: 1) the h=8 variants used to estimate the initial LTS estimate $\hat{\theta}_{LTS,h}$; 2) the variants with a weight of 1 in $w_{LTS,2}$; and 3) the variants selected from the automated approach based on the heterogeneity stopping criteria.

Appendix I

Chapter 6 supplementary material

I.1 One–sample Mendelian randomization study using data from UK Biobank

I.1.1 Histograms of the adiposity and body composition measurements

Histograms of weight, height and BMI are contained in Figure I.1, and Figure I.2 contains histograms of the BIA measurements for FMI and FFMI. Figures I.1 and I.2 consist of the data used in the main analysis (n=360,409), and are presented for the entire dataset, and by gender. Figure I.3 contains histograms of the DXA measurements for FMI, FFMI and LMI, and these histograms are based on the 3,901 participants with DXA measurements.

Fig. I.1 Histograms of weight, height and body mass index (BMI) by gender of the 360,409 participants in UK Biobank included in the main analysis for the entire dataset and by gender. The dotted red lines represent the mean values of the measurements.

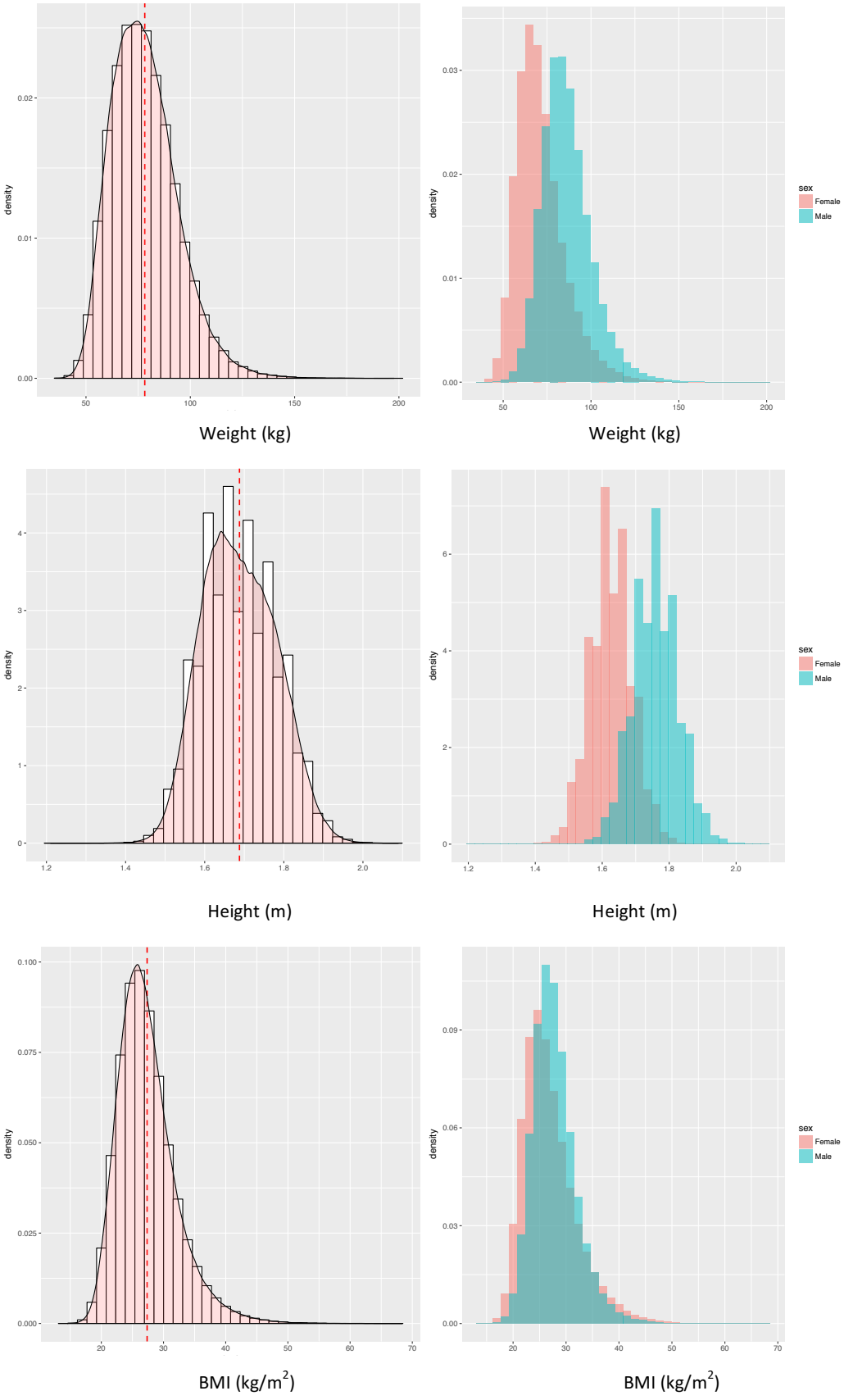


Fig. I.2 Histograms of the bioelectrical impedance analysis measurements of fat mass index (FMI) and fat-free mass index (FFMI) by gender of the 360,409 participants in UK Biobank included in the main analysis for the entire dataset and by gender. The dotted red lines represent the mean values of the measurements.

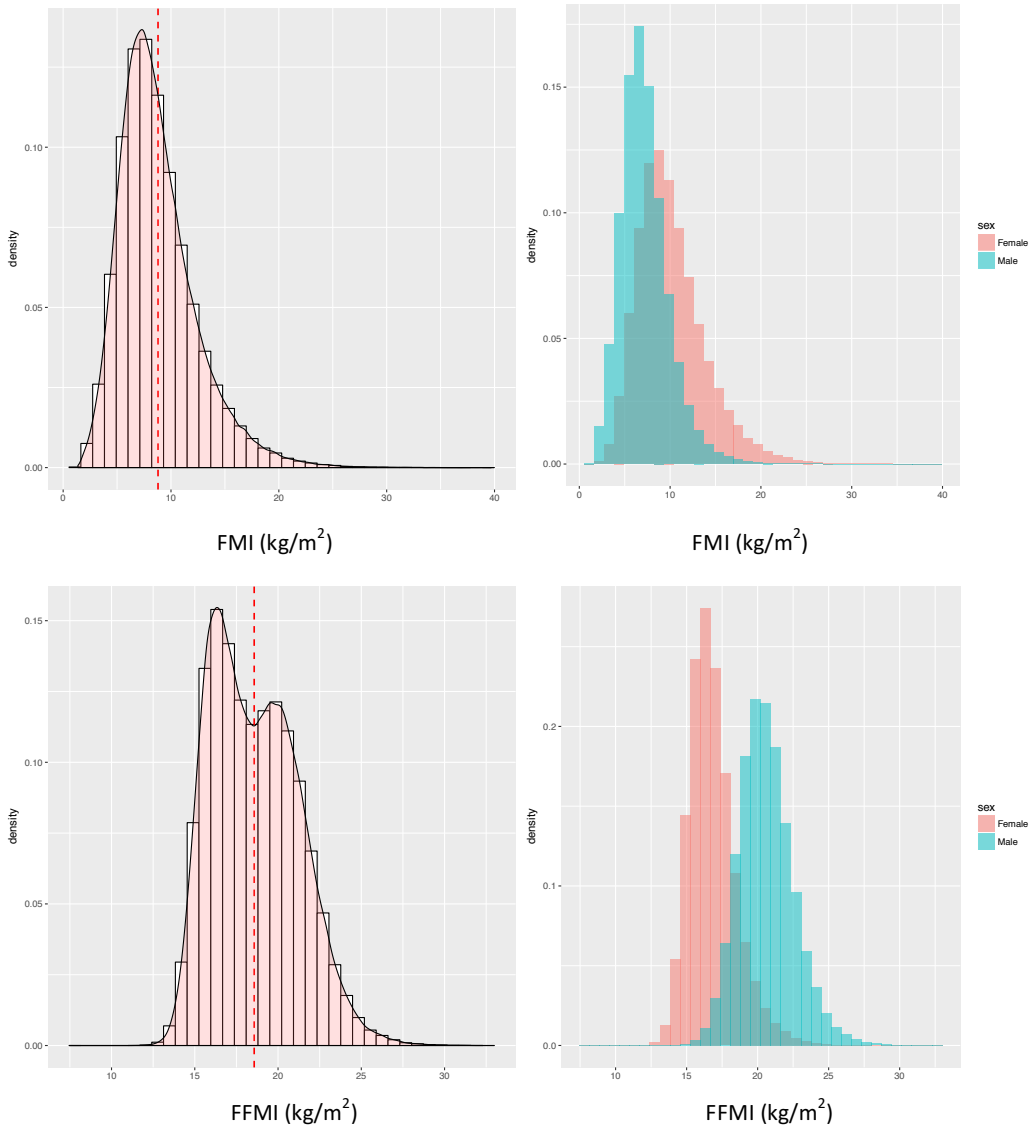
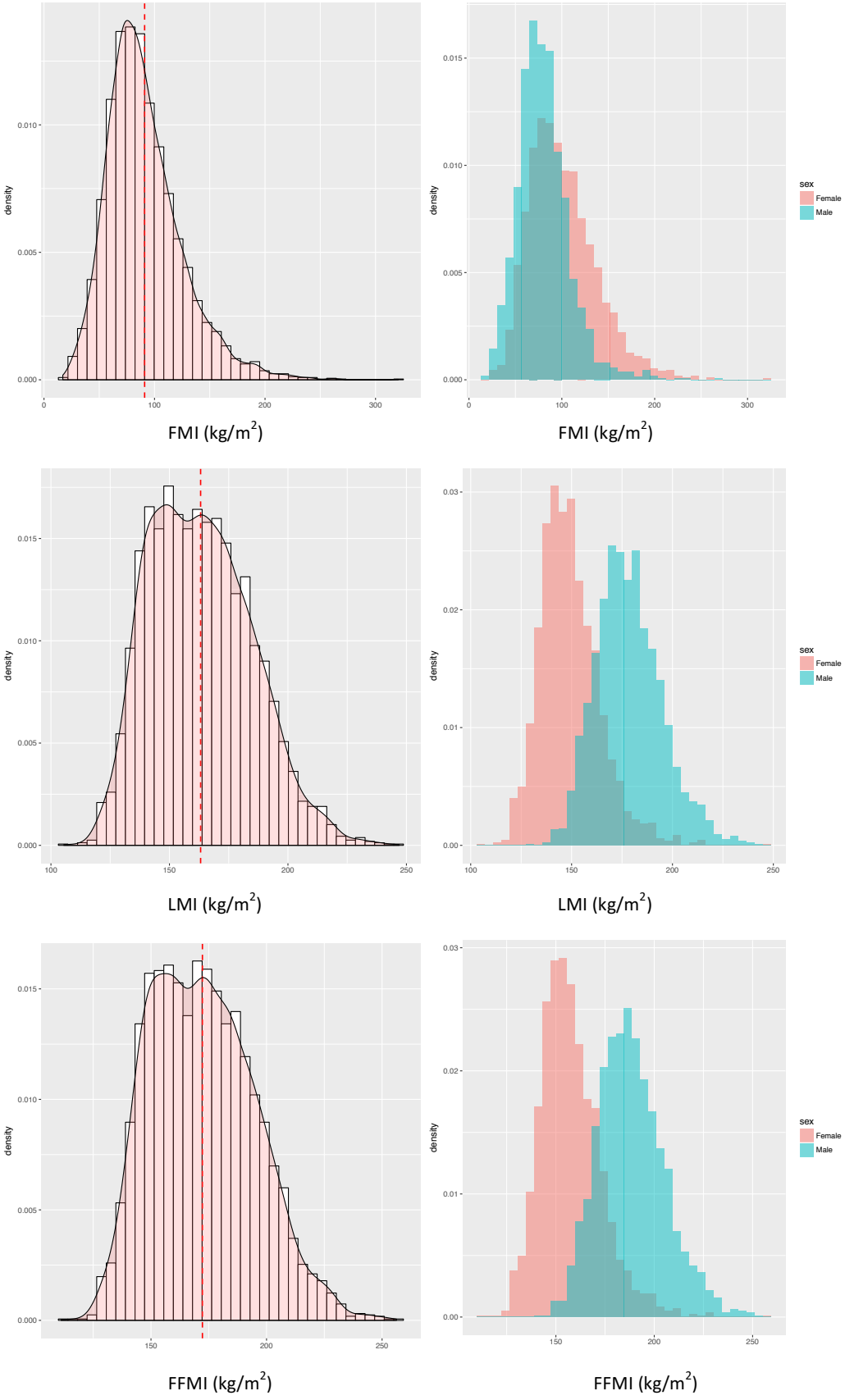


Fig. I.3 Histograms of the dual-energy X-ray (DXA) measurements fat mass index (FMI), fat-free mass index (FFMI) and lean mass index (LMI) of the 3,901 participants with DXA measurements by gender. The dotted red lines represent the mean values of the measurements.



I.1.2 Information on the genetic variants considered for the Mendelian randomization analysis

Table I.1 contains information on the genetic variants considered for the Mendelian randomization analysis in adults using the UK Biobank dataset. The final two columns in Table I.1 indicate whether the variant was included in the liberal or conservative sets of genetic variants.

Table I.1 Information on the 77 genetic variants selected from the GIANT consortium [104] and were considered for the Mendelian randomization analysis using the UK Biobank dataset. The final two columns indicate whether the variant was included in the liberal or conservative sets of genetic variants.

rs no.	Chr.	Gene	Minor/ major allele	MAF	Imputation quality	Selection criteria	
						Liberal	Con.
rs11165643	1	*PTBP2*	C/T	0.4091	0.9981	Yes	Yes
rs11583200	1	*ELAVL4*	C/T	0.3872	0.9939	No	No
rs12401738	1	*FUBP1*	A/G	0.3793	0.9958	Yes	No
rs12566985	1	*FPGT-TNNI3K*	G/A	0.4376	0.9985	Yes	Yes
rs17024393	1	*GNAT2*	C/T	0.0260	0.9909	Yes	Yes
rs2820292	1	*NAV1*	A/C	0.4339	1.0000	Yes	No
rs3101336	1	*NEGR1*	T/C	0.3974	1.0000	No	No
rs543874	1	*SEC16B*	G/A	0.2058	1.0000	Yes	No
rs657452	1	*AGBL4*	A/G	0.3913	0.9887	Yes	Yes
rs1016287	2	*LINC01122*	T/C	0.2983	0.9979	Yes	No
rs10182181	2	*ADCY3*	G/A	0.4867	0.9966	No	No
rs11126666	2	*KCNK3*	A/G	0.2555	0.9970	Yes	No
rs11688816	2	*EHBP1*	A/G	0.4593	0.9936	Yes	Yes
rs13021737	2	*TMEM18*	A/G	0.1714	0.9999	No	No
rs1528435	2	*UBE2E3*	C/T	0.3791	0.9978	Yes	No
rs2121279	2	*LRP1B*	T/C	0.1255	0.9918	No	No
rs7599312	2	*ERBB4*	A/G	0.2677	0.9811	Yes	Yes
rs13078960	3	*CADM2*	G/T	0.2009	0.9941	Yes	Yes
rs1516725	3	*ETV5*	T/C	0.1371	0.9958	Yes	No
rs16851483	3	*RASA2*	T/G	0.0660	0.9991	Yes	No
rs2365389	3	*FHIT*	T/C	0.4083	0.9948	Yes	No
rs3849570	3	*GBE1*	A/C	0.3466	0.9998	Yes	Yes
rs6804842	3	*RARB*	A/G	0.4267	0.9930	Yes	Yes

rs10938397	4	*GNPDA2*	G/A	0.4344	1.0000	Yes	No
rs11727676	4	*HHIP*	C/T	0.0962	1.0000	Yes	Yes
rs13107325	4	*SLC39A8*	T/C	0.0747	1.0000	No	No
rs17001654	4	*NUP54*	G/C	0.1485	0.9760	Yes	Yes
rs2112347	5	*POC5*	G/T	0.3598	1.0000	Yes	No
rs13191362	6	*PARK2*	G/A	0.1246	0.9953	Yes	No
rs2033529	6	*TDRG1*	G/A	0.2882	0.9922	Yes	Yes
rs205262	6	*C6orf106*	G/A	0.2683	0.9986	Yes	No
rs2207139	6	*TFAP2B*	G/A	0.1696	0.9997	Yes	Yes
rs9400239	6	*FOXO3*	T/C	0.2952	0.9957	No	No
rs1167827	7	*HIP1*	A/G	0.4349	1.0000	Yes	No
rs2245368	7	*PMS2L11*	C/T	0.1686	1.0000	Yes	No
rs17405819	8	*HNF4G*	C/T	0.2980	0.9999	Yes	No
rs2033732	8	*RALYL*	T/C	0.2551	1.0000	Yes	Yes
rs10733682	9	*LMX1B*	A/G	0.4733	0.9664	Yes	Yes
rs10968576	9	*LINGO2*	G/A	0.3222	1.0000	Yes	Yes
rs1928295	9	*TLR4*	C/T	0.4306	0.9999	Yes	No
rs4740619	9	*C9orf93*	C/T	0.4490	0.9990	Yes	No
rs6477694	9	*EPB41L4B*	C/T	0.3531	0.9919	Yes	Yes
rs11191560	10	*NT5C2*	C/T	0.0776	0.9998	Yes	Yes
rs17094222	10	*HIF1AN*	C/T	0.2131	0.9922	Yes	Yes
rs7899106	10	*GRID1*	G/A	0.0501	0.9905	Yes	Yes
rs7903146	10	*TCF7L2*	T/C	0.2906	1.0000	Yes	Yes
rs11030104	11	*BDNF*	G/A	0.2031	0.9990	No	No
rs12286929	11	*CADM1*	A/G	0.4734	0.9969	Yes	No
rs2176598	11	*HSD17B12*	T/C	0.2466	1.0000	Yes	Yes
rs3817334	11	*MTCH2*	T/C	0.4079	1.0000	Yes	Yes
rs4256980	11	*TRIM66*	C/G	0.3458	0.9963	Yes	Yes
rs11057405	12	*CLIP1*	A/G	0.1051	1.0000	Yes	Yes
rs7138803	12	*BCDIN3D*	A/G	0.3688	1.0000	Yes	Yes
rs12429545	13	*OLFM4*	A/G	0.1292	0.9820	Yes	Yes
rs9581854	13	*MTIF3*	T/C	0.1796	0.9968	No	No
rs10132280	14	*STXBP6*	A/C	0.3005	0.9891	Yes	Yes
rs11847697	14	*PRKD1*	T/C	0.0437	1.0000	Yes	Yes
rs12885454	14	*PRKD1*	A/C	0.3561	0.9984	Yes	Yes
rs7141420	14	*NRXN3*	C/T	0.4835	0.9883	Yes	No
rs16951275	15	*MAP2K5*	C/T	0.2257	0.9994	No	No
rs3736485	15	*SCG3*	A/G	0.4609	0.9953	Yes	Yes
rs12446632	16	*GPRC5B*	A/G	0.1425	0.9998	Yes	Yes

rs1558902	16	*FTO*	A/T	0.4031	0.9998	Yes	Yes
rs2650492	16	*SBK1*	A/G	0.2971	0.9895	No	No
rs3888190	16	*ATP2A1*	A/C	0.3997	0.9999	No	No
rs758747	16	*NLRC3*	T/C	0.2780	0.9791	Yes	No
rs9925964	16	*KAT8*	G/A	0.3587	0.9982	Yes	Yes
rs1000940	17	*RABEP1*	G/A	0.3015	0.9987	Yes	Yes
rs12940622	17	*RPTOR*	A/G	0.4400	0.9990	Yes	Yes
rs1808579	18	*C18orf8*	T/C	0.4832	0.9982	No	No
rs6567160	18	*MC4R*	C/T	0.2329	0.9988	Yes	Yes
rs7243357	18	*GRP*	G/T	0.1769	0.9919	Yes	Yes
rs17724992	19	*GDF15*	G/A	0.2677	0.9917	No	No
rs2075650	19	*TOMM40*	G/A	0.1458	1.0000	No	No
rs2287019	19	*QPCTL*	T/C	0.1824	0.9856	No	No
rs29941	19	*KCTD15*	A/G	0.3262	1.0000	No	No
rs3810291	19	*ZC3H4*	G/A	0.3247	1.0000	Yes	Yes

Abbreviations: no., number; Chr., chromosome; MAF, minor allele frequency; Con., conservative.

I.1.3 Genetic associations with the body composition measurements and asthma

Tables I.2 to I.4 and Table I.7 contain estimates of the genetic associations for BMI, FMI (BIA measurements), FFMI (BIA measurements) and asthma (ever diagnosis and current) based on the 360,409 participants from UK Biobank used in the main analysis. Tables I.5 to I.6 contain estimates of the genetic associations of the DXA measurements for FMI and FFMI for the 3,901 participants with data on these measurements. The genetic associations were adjusted for the first 10 PCs and gender, and the adiposity associations were also adjusted for height.

Table I.2 Estimates (standard errors) and p-values of the genetic associations with body mass index for the 60 genetic variants selected under the liberal criteria based on 360,409 participants from UK Biobank. The R^2 value and F-statistic are provided from the model when the body composition measurement was regressed against the genetic variant.

rs no.	Chr.	Gene	Alleles[a]	Estimate (se)	P-value	R^2	F-stat
rs11165643	1	*PTBP2*	T/C	0.086 (0.011)	2.32×10^{-14}	0.00017	60.0
rs12401738	1	*FUBP1*	A/G	0.091 (0.011)	1.72×10^{-15}	0.00019	68.4
rs12566985	1	*FPGT-TNNI3K*	G/A	0.090 (0.011)	6.70×10^{-16}	0.00018	63.4
rs17024393	1	*GNAT2*	C/T	0.331 (0.035)	2.01×10^{-21}	0.00023	81.6
rs2820292	1	*NAV1*	C/A	0.091 (0.011)	3.55×10^{-16}	0.00019	70.1
rs543874	1	*SEC16B*	G/A	0.245 (0.014)	8.41×10^{-72}	0.00089	322.0
rs657452	1	*AGBL4*	A/G	0.080 (0.011)	1.60×10^{-12}	0.00013	48.6
rs1016287	2	*LINC01122*	T/C	0.088 (0.012)	3.44×10^{-13}	0.00015	55.9
rs11126666	2	*KCNK3*	A/G	0.018 (0.013)	0.15529	0.00000	1.5
rs11688816	2	*EHBP1*	G/A	0.049 (0.011)	0.00001	0.00006	22.7
rs1528435	2	*UBE2E3*	T/C	0.075 (0.011)	4.28×10^{-11}	0.00011	40.3
rs7599312	2	*ERBB4*	G/A	0.077 (0.013)	8.00×10^{-10}	0.00012	43.3
rs13078960	3	*CADM2*	G/T	0.086 (0.014)	3.91×10^{-10}	0.00010	34.8
rs1516725	3	*ETV5*	C/T	0.148 (0.016)	2.40×10^{-20}	0.00022	79.8
rs16851483	3	*RASA2*	T/G	0.163 (0.022)	2.33×10^{-13}	0.00017	61.2
rs2365389	3	*FHIT*	C/T	0.082 (0.011)	3.89×10^{-13}	0.00014	51.1
rs3849570	3	*GBE1*	A/C	0.048 (0.012)	0.00004	0.00005	18.7
rs6804842	3	*RARB*	G/A	0.068 (0.011)	9.80×10^{-10}	0.00008	29.7
rs10938397	4	*GNPDA2*	G/A	0.147 (0.011)	1.02×10^{-39}	0.00051	184.0
rs11727676	4	*HHIP*	T/C	0.018 (0.019)	0.34962	0.00001	2.5
rs17001654	4	*NUP54*	G/C	0.069 (0.016)	0.00001	0.00004	15.9

rs2112347	5	*POC5*	T/G	0.132 (0.012)	1.37×10^{-30}	0.00035	126.4
rs13191362	6	*PARK2*	A/G	0.083 (0.017)	7.37×10^{-7}	0.00007	25.7
rs2033529	6	*TDRG1*	G/A	0.098 (0.012)	1.20×10^{-15}	0.00017	62.4
rs205262	6	*C6orf106*	G/A	0.137 (0.012)	3.25×10^{-28}	0.00028	100.3
rs2207139	6	*TFAP2B*	G/A	0.186 (0.015)	9.31×10^{-37}	0.00047	168.1
rs1167827	7	*HIP1*	G/A	0.102 (0.011)	7.06×10^{-20}	0.00025	90.8
rs2245368	7	*PMS2L11*	C/T	0.122 (0.015)	1.08×10^{-16}	0.00015	54.6
rs17405819	8	*HNF4G*	T/C	0.102 (0.012)	3.58×10^{-17}	0.00019	69.3
rs2033732	8	*RALYL*	C/T	0.049 (0.013)	0.00011	0.00004	13.5
rs10733682	9	*LMX1B*	A/G	0.063 (0.011)	1.71×10^{-8}	0.00009	32.4
rs10968576	9	*LINGO2*	G/A	0.112 (0.012)	2.38×10^{-21}	0.00029	106.1
rs1928295	9	*TLR4*	T/C	0.058 (0.011)	2.41×10^{-7}	0.00008	29.5
rs4740619	9	*C9orf93*	T/C	0.085 (0.011)	2.26×10^{-14}	0.00017	62.2
rs6477694	9	*EPB41L4B*	C/T	0.062 (0.012)	9.54×10^{-8}	0.00006	20.8
rs11191560	10	*NT5C2*	C/T	0.122 (0.021)	3.49×10^{-9}	0.00009	31.3
rs17094222	10	*HIF1AN*	C/T	0.066 (0.014)	8.77×10^{-7}	0.00007	26.6
rs7899106	10	*GRID1*	G/A	0.141 (0.025)	3.05×10^{-8}	0.00007	26.4
rs7903146	10	*TCF7L2*	C/T	0.075 (0.012)	5.45×10^{-10}	0.00011	38.7
rs12286929	11	*CADM1*	G/A	0.062 (0.011)	1.97×10^{-8}	0.00008	30.6
rs2176598	11	*HSD17B12*	T/C	0.091 (0.013)	1.58×10^{-12}	0.00014	49.6
rs3817334	11	*MTCH2*	T/C	0.106 (0.011)	3.32×10^{-21}	0.00031	112.8
rs4256980	11	*TRIM66*	G/C	0.074 (0.012)	1.52×10^{-10}	0.00010	37.4
rs11057405	12	*CLIP1*	G/A	0.150 (0.018)	1.04×10^{-16}	0.00014	49.3
rs7138803	12	*BCDIN3D*	A/G	0.127 (0.011)	8.96×10^{-29}	0.00031	112.8
rs12429545	13	*OLFM4*	A/G	0.137 (0.017)	1.61×10^{-16}	0.00017	61.3
rs10132280	14	*STXBP6*	C/A	0.101 (0.012)	8.58×10^{-17}	0.00021	74.3
rs11847697	14	*PRKD1*	T/C	0.108 (0.027)	0.00006	0.00004	13.4
rs12885454	14	*PRKD1*	C/A	0.079 (0.012)	8.88×10^{-12}	0.00013	47.7
rs7141420	14	*NRXN3*	T/C	0.098 (0.011)	1.51×10^{-18}	0.00020	72.8
rs3736485	15	*SCG3*	A/G	0.056 (0.011)	4.31×10^{-7}	0.00006	22.8
rs12446632	16	*GPRC5B*	G/A	0.138 (0.016)	2.58×10^{-18}	0.00020	71.0
rs1558902	16	*FTO*	A/T	0.353 (0.011)	9.29×10^{-217}	0.00268	968.8
rs758747	16	*NLRC3*	T/C	0.052 (0.012)	0.00003	0.00005	18.1
rs9925964	16	*KAT8*	A/G	0.127 (0.012)	2.94×10^{-28}	0.00036	128.6
rs1000940	17	*RABEP1*	G/A	0.070 (0.012)	4.76×10^{-9}	0.00009	31.2
rs12940622	17	*RPTOR*	G/A	0.096 (0.011)	4.33×10^{-18}	0.00020	72.5
rs6567160	18	*MC4R*	C/T	0.279 (0.013)	2.05×10^{-101}	0.00105	380.6
rs7243357	18	*GRP*	T/G	0.100 (0.015)	6.91×10^{-12}	0.00011	38.4

rs3810291	19	*ZC3H4*	A/G	0.128 (0.012)	2.43×10^{-27}	0.00031	110.5

Abbreviations: no., number; Chr., chromosome; se, standard error; F-stat, F-statistic.
[a]Allele increasing/allele decreasing: genetic associations reported with respect to the trait increasing allele.

Table I.3 Estimates (standard errors) and p-values of the genetic associations with the bioelectrical impedance analysis (BIA) measurements for fat mass index for the 60 genetic variants selected under the liberal criteria based on 360,409 participants from UK Biobank. The R^2 value and F-statistic are provided from the model when the body composition measurement was regressed against the genetic variant.

rs no.	Chr.	Gene	Alleles[a]	Estimate (se)	P-value	R^2	F-stat
rs11165643	1	*PTBP2*	T/C	0.061 (0.008)	7.37×10^{-15}	0.00016	56.5
rs12401738	1	*FUBP1*	A/G	0.054 (0.008)	9.68×10^{-12}	0.00015	54.4
rs12566985	1	*FPGT*	G/A	0.053 (0.008)	7.00×10^{-12}	0.00011	38
		TNNI3K					
rs17024393	1	*GNAT2*	C/T	0.221 (0.024)	1.29×10^{-19}	0.00020	71.6
rs2820292	1	*NAV1*	C/A	0.058 (0.008)	1.15×10^{-13}	0.00016	57.1
rs543874	1	*SEC16B*	G/A	0.157 (0.010)	1.40×10^{-60}	0.00065	235.2
rs657452	1	*AGBL4*	A/G	0.056 (0.008)	2.34×10^{-12}	0.00011	38.4
rs1016287	2	*LINC01122*	T/C	0.059 (0.008)	2.96×10^{-12}	0.00009	34.1
rs11126666	2	*KCNK3*	A/G	0.004 (0.009)	0.612284	0.00000	0.0
rs11688816	2	*EHBP1*	G/A	0.031 (0.008)	0.00007	0.00005	14.4
rs1528435	2	*UBE2E3*	T/C	0.054 (0.008)	8.88×10^{-12}	0.00010	35.1
rs7599312	2	*ERBB4*	G/A	0.045 (0.009)	3.09×10^{-07}	0.00006	22.6
rs13078960	3	*CADM2*	G/T	0.048 (0.010)	7.02×10^{-07}	0.00006	21.1
rs1516725	3	*ETV5*	C/T	0.082 (0.011)	2.41×10^{-13}	0.00010	36.4
rs16851483	3	*RASA2*	T/G	0.109 (0.016)	2.32×10^{-12}	0.00015	52.8
rs2365389	3	*FHIT*	C/T	0.049 (0.008)	3.39×10^{-10}	0.00010	36.6
rs3849570	3	*GBE1*	A/C	0.036 (0.008)	8.38×10^{-06}	0.00005	19.3
rs6804842	3	*RARB*	G/A	0.038 (0.008)	1.10×10^{-06}	0.00003	10.6
rs10938397	4	*GNPDA2*	G/A	0.100 (0.008)	7.39×10^{-38}	0.00032	115.5
rs11727676	4	*HHIP*	C/T	0.019 (0.013)	0.148552	0.00000	0.5
rs17001654	4	*NUP54*	G/C	0.040 (0.011)	0.00032	0.00002	6.9
rs2112347	5	*POC5*	T/G	0.090 (0.008)	4.16×10^{-29}	0.00026	94.4
rs13191362	6	*PARK2*	A/G	0.045 (0.012)	0.00012	0.00003	9.3
rs2033529	6	*TDRG1*	G/A	0.060 (0.009)	1.52×10^{-12}	0.00013	48.5
rs205262	6	*C6orf106*	G/A	0.090 (0.009)	3.12×10^{-25}	0.00017	62.1
rs2207139	6	*TFAP2B*	G/A	0.116 (0.010)	1.32×10^{-29}	0.00032	117.1
rs1167827	7	*HIP1*	G/A	0.064 (0.008)	2.19×10^{-16}	0.00013	45.6
rs2245368	7	*PMS2L11*	C/T	0.068 (0.010)	4.73×10^{-11}	0.00011	38.7
rs17405819	8	*HNF4G*	T/C	0.071 (0.008)	3.88×10^{-17}	0.00016	56.6
rs2033732	8	*RALYL*	C/T	0.025 (0.009)	0.00446	0.00001	4
rs10733682	9	*LMX1B*	A/G	0.038 (0.008)	1.69×10^{-6}	0.00005	17.4
rs10968576	9	*LINGO2*	G/A	0.069 (0.008)	3.59×10^{-17}	0.00028	100.1

rs1928295	9	*TLR4*	T/C	0.039 (0.008)	5.44×10^{-7}	0.00006	20.6
rs4740619	9	*C9orf93*	T/C	0.068 (0.008)	2.95×10^{-18}	0.00020	72.1
rs6477694	9	*EPB41L4B*	C/T	0.038 (0.008)	3.33×10^{-6}	0.00005	17.5
rs11191560	10	*NT5C2*	C/T	0.066 (0.014)	4.79×10^{-6}	0.00003	10.2
rs17094222	10	*HIF1AN*	C/T	0.048 (0.009)	4.21×10^{-7}	0.00008	29.6
rs7899106	10	*GRID1*	G/A	0.079 (0.018)	7.86×10^{-6}	0.00003	12.4
rs7903146	10	*TCF7L2*	C/T	0.050 (0.008)	2.93×10^{-9}	0.00010	35.3
rs12286929	11	*CADM1*	G/A	0.037 (0.008)	2.00×10^{-6}	0.00005	18.7
rs2176598	11	*HSD17B12*	T/C	0.058 (0.009)	7.00×10^{-11}	0.00011	40.9
rs3817334	11	*MTCH2*	T/C	0.086 (0.008)	8.30×10^{-28}	0.00036	129.5
rs4256980	11	*TRIM66*	G/C	0.053 (0.008)	7.52×10^{-11}	0.00013	45.2
rs11057405	12	*CLIP1*	G/A	0.113 (0.013)	3.82×10^{-19}	0.00014	48.8
rs7138803	12	*BCDIN3D*	A/G	0.084 (0.008)	4.84×10^{-26}	0.00025	89.1
rs12429545	13	*OLFM4*	A/G	0.088 (0.012)	3.28×10^{-14}	0.00010	36.2
rs10132280	14	*STXBP6*	C/A	0.053 (0.008)	3.19×10^{-10}	0.00012	42.8
rs11847697	14	*PRKD1*	T/C	0.071 (0.019)	0.00018	0.00003	11.5
rs12885454	14	*PRKD1*	C/A	0.050 (0.008)	5.11×10^{-10}	0.00009	32.2
rs7141420	14	*NRXN3*	T/C	0.065 (0.008)	4.11×10^{-17}	0.00013	46.7
rs3736485	15	*SCG3*	A/G	0.044 (0.008)	1.35×10^{-8}	0.00006	23.3
rs12446632	16	*GPRC5B*	G/A	0.069 (0.011)	4.85×10^{-10}	0.00006	20.5
rs1558902	16	*FTO*	A/T	0.217 (0.008)	7.79×10^{-168}	0.00145	523.6
rs758747	16	*NLRC3*	T/C	0.032 (0.009)	0.00023	0.00004	12.8
rs9925964	16	*KAT8*	A/G	0.089 (0.008)	1.58×10^{-28}	0.00037	135.1
rs1000940	17	*RABEP1*	G/A	0.048 (0.008)	1.22×10^{-8}	0.00009	31.5
rs12940622	17	*RPTOR*	G/A	0.063 (0.008)	6.52×10^{-16}	0.00015	53.9
rs6567160	18	*MC4R*	C/T	0.161 (0.009)	1.90×10^{-69}	0.00062	225.3
rs7243357	18	*GRP*	T/G	0.061 (0.010)	1.89×10^{-9}	0.00007	25.2
rs3810291	19	*ZC3H4*	A/G	0.067 (0.008)	2.90×10^{-16}	0.00016	58.8

Abbreviations: no., number; Chr., chromosome; se, standard error; F-stat, F-statistic.
[a]Allele increasing/allele decreasing: genetic associations reported with respect to the trait increasing allele.

Table I.4 Estimates (standard errors) and p-values of the genetic associations with the bioelectrical impedance analysis (BIA) measurements for fat-free mass index for the 60 genetic variants selected under the liberal criteria based on 360,409 participants from UK Biobank. The R^2 value and F-statistic are provided from the model when the body composition measurement was regressed against the genetic variant.

rs no.	Chr.	Gene	Alleles[a]	Estimate (se)	P-value	R^2	F-stat
rs11165643	1	*PTBP2*	T/C	0.025 (0.004)	8.59×10^{-9}	0.00004	13.4
rs12401738	1	*FUBP1*	A/G	0.037 (0.004)	4.11×10^{-17}	0.00006	23.2
rs12566985	1	*FPGT*	G/A	0.037 (0.004)	6.25×10^{-18}	0.00010	35.1
		TNNI3K					
rs17024393	1	*GNAT2*	C/T	0.111 (0.013)	7.93×10^{-17}	0.00006	22.1
rs2820292	1	*NAV1*	C/A	0.033 (0.004)	7.06×10^{-15}	0.00006	22.6
rs543874	1	*SEC16B*	G/A	0.088 (0.005)	2.24×10^{-64}	0.00036	129.3
rs657452	1	*AGBL4*	A/G	0.025 (0.004)	1.39×10^{-8}	0.00005	16.6
rs1016287	2	*LINC01122*	T/C	0.029 (0.005)	3.05×10^{-10}	0.00008	30.1
rs11126666	2	*KCNK3*	A/G	0.013 (0.005)	0.00524	0.00001	4.0
rs11688816	2	*EHBP1*	G/A	0.018 (0.004)	0.00003	0.00003	11.5
rs1528435	2	*UBE2E3*	T/C	0.021 (0.004)	1.57×10^{-06}	0.00003	11.1
rs7599312	2	*ERBB4*	G/A	0.032 (0.005)	1.46×10^{-11}	0.00008	29.0
rs13078960	3	*CADM2*	G/T	0.039 (0.005)	2.51×10^{-13}	0.00005	18.8
rs1516725	3	*ETV5*	C/T	0.066 (0.006)	2.17×10^{-27}	0.00017	62.0
rs16851483	3	*RASA2*	T/G	0.054 (0.008)	1.74×10^{-10}	0.00005	17.3
rs2365389	3	*FHIT*	C/T	0.032 (0.004)	4.72×10^{-14}	0.00006	21.3
rs3849570	3	*GBE1*	A/C	0.012 (0.004)	0.00787	0.00001	3.2
rs6804842	3	*RARB*	G/A	0.030 (0.004)	1.25×10^{-12}	0.00008	29.0
rs10938397	4	*GNPDA2*	G/A	0.047 (0.004)	2.43×10^{-28}	0.00026	95.3
rs11727676	4	*HHIP*	T/C	0.036 (0.007)	3.43×10^{-7}	0.00004	15.0
rs17001654	4	*NUP54*	G/C	0.029 (0.006)	1.20×10^{-6}	0.00004	13.0
rs2112347	5	*POC5*	T/G	0.042 (0.004)	5.29×10^{-22}	0.00013	48.6
rs13191362	6	*PARK2*	A/G	0.038 (0.006)	2.50×10^{-9}	0.00007	24.7
rs2033529	6	*TDRG1*	G/A	0.038 (0.005)	7.78×10^{-16}	0.00006	22.3
rs205262	6	*C6orf106*	G/A	0.047 (0.005)	4.96×10^{-23}	0.00015	53.0
rs2207139	6	*TFAP2B*	G/A	0.070 (0.006)	6.43×10^{-36}	0.00020	73.3
rs1167827	7	*HIP1*	G/A	0.038 (0.004)	6.59×10^{-19}	0.00018	63.1
rs2245368	7	*PMS2L11*	C/T	0.055 (0.006)	2.69×10^{-22}	0.00006	23.1
rs17405819	8	*HNF4G*	T/C	0.031 (0.005)	2.14×10^{-11}	0.00006	22.3
rs2033732	8	*RALYL*	C/T	0.024 (0.005)	1.01×10^{-6}	0.00004	15.2
rs10733682	9	*LMX1B*	A/G	0.026 (0.004)	1.49×10^{-9}	0.00006	21.0
rs10968576	9	*LINGO2*	G/A	0.042 (0.005)	3.76×10^{-21}	0.00007	23.6

rs1928295	9	*TLR4*	T/C	0.019 (0.004)	0.00002	0.00004	12.8
rs4740619	9	*C9orf93*	T/C	0.017 (0.004)	0.00004	0.00002	6.7
rs6477694	9	*EPB41L4B*	C/T	0.024 (0.004)	4.09×10^{-8}	0.00002	6.3
rs11191560	10	*NT5C2*	C/T	0.056 (0.008)	1.16×10^{-12}	0.00009	33.0
rs17094222	10	*HIF1AN*	C/T	0.019 (0.005)	0.00029	0.00001	3.4
rs7899106	10	*GRID1*	G/A	0.061 (0.010)	2.33×10^{-10}	0.00006	19.9
rs7903146	10	*TCF7L2*	C/T	0.025 (0.005)	6.95×10^{-8}	0.00003	9.4
rs12286929	11	*CADM1*	G/A	0.025 (0.004)	1.70×10^{-9}	0.00005	16.6
rs2176598	11	*HSD17B12*	T/C	0.032 (0.005)	4.40×10^{-11}	0.00004	15.6
rs3817334	11	*MTCH2*	T/C	0.020 (0.004)	1.77×10^{-6}	0.00003	12.6
rs4256980	11	*TRIM66*	G/C	0.022 (0.004)	1.10×10^{-6}	0.00001	3.3
rs11057405	12	*CLIP1*	G/A	0.037 (0.007)	6.87×10^{-8}	0.00003	9.5
rs7138803	12	*BCDIN3D*	A/G	0.043 (0.004)	5.76×10^{-23}	0.00011	38.8
rs12429545	13	*OLFM4*	A/G	0.049 (0.006)	1.06×10^{-14}	0.00010	34.6
rs10132280	14	*STXBP6*	C/A	0.048 (0.005)	6.98×10^{-25}	0.00012	43.5
rs11847697	14	*PRKD1*	T/C	0.038 (0.010)	0.00028	0.00001	3.8
rs12885454	14	*PRKD1*	C/A	0.029 (0.004)	7.69×10^{-11}	0.00006	22.0
rs7141420	14	*NRXN3*	T/C	0.032 (0.004)	1.87×10^{-14}	0.00040	36.5
rs3736485	15	*SCG3*	A/G	0.012 (0.004)	0.00436	0.00001	4
rs12446632	16	*GPRC5B*	G/A	0.069 (0.006)	1.37×10^{-30}	0.00023	81.5
rs1558902	16	*FTO*	A/T	0.137 (0.004)	2.06×10^{-222}	0.00171	616.0
rs758747	16	*NLRC3*	T/C	0.020 (0.005)	0.00003	0.00002	7.6
rs9925964	16	*KAT8*	A/G	0.038 (0.004)	6.73×10^{-18}	0.00006	20.3
rs1000940	17	*RABEP1*	G/A	0.023 (0.005)	8.71×10^{-7}	0.00002	5.7
rs12940622	17	*RPTOR*	G/A	0.034 (0.004)	1.87×10^{-15}	0.00008	28.3
rs6567160	18	*MC4R*	C/T	0.119 (0.005)	3.30×10^{-125}	0.00059	214.3
rs7243357	18	*GRP*	T/G	0.039 (0.006)	2.53×10^{-12}	0.00005	18.6
rs3810291	19	*ZC3H4*	A/G	0.060 (0.004)	3.38×10^{-41}	0.00020	71.9

Abbreviations: no., number; Chr., chromosome; se, standard error; F-stat, F-statistic.
[a]Allele increasing/allele decreasing: genetic associations reported with respect to the trait increasing allele.

Table I.5 Estimates (standard errors) and p-values of the genetic associations with the dual-energy X-ray (DXA) measurements for fat mass index for the 60 genetic variants selected under the liberal criteria based on 3,901 participants from UK Biobank. The R^2 value and F-statistic are provided from the model when the body composition measurement was regressed against the genetic variant.

rs no.	Chr.	Gene	Alleles[a]	Estimate (se)	P-value	R^2	F-stat
rs11165643	1	*PTBP2*	C/T	0.038 (0.075)	0.61282	0.00008	0.3
rs12401738	1	*FUBP1*	A/G	0.106 (0.075)	0.15896	0.00084	3.3
rs12566985	1	*FPGT*	A/G	0.028 (0.075)	0.71242	0.00005	0.2
		TNNI3K					
rs17024393	1	*GNAT2*	C/T	0.456 (0.246)	0.06379	0.00091	3.5
rs2820292	1	*NAV1*	C/A	0.039 (0.074)	0.59614	0.00022	0.9
rs543874	1	*SEC16B*	G/A	0.064 (0.091)	0.48133	0.00004	0.2
rs657452	1	*AGBL4*	G/A	0.051 (0.075)	0.49583	0.00023	0.9
rs1016287	2	*LINC01122*	C/T	0.080 (0.080)	0.31698	0.00027	1.1
rs11126666	2	*KCNK3*	G/A	0.071 (0.084)	0.39654	0.00011	0.4
rs11688816	2	*EHBP1*	A/G	0.009 (0.074)	0.90260	0.00001	0.0
rs1528435	2	*UBE2E3*	T/C	0.036 (0.075)	0.62743	0.00000	0.0
rs7599312	2	*ERBB4*	G/A	0.077 (0.083)	0.35152	0.00045	1.7
rs13078960	3	*CADM2*	G/T	0.022 (0.092)	0.8148	0.00005	0.2
rs1516725	3	*ETV5*	T/C	0.048 (0.107)	0.65129	0.00025	1.0
rs16851483	3	*RASA2*	G/T	0.100 (0.149)	0.50337	0.00007	0.3
rs2365389	3	*FHIT*	C/T	0.091 (0.075)	0.22437	0.00089	3.5
rs3849570	3	*GBE1*	C/A	0.061 (0.076)	0.42421	0.00004	0.2
rs6804842	3	*RARB*	A/G	0.015 (0.074)	0.83936	0.00007	0.3
rs10938397	4	*GNPDA2*	A/G	0.003 (0.075)	0.96626	0.00006	0.2
rs11727676	4	*HHIP*	T/C	0.010 (0.123)	0.93463	0.00006	0.2
rs17001654	4	*NUP54*	G/C	0.143 (0.103)	0.16758	0.00015	0.6
rs2112347	5	*POC5*	G/T	0.113 (0.076)	0.13553	0.00042	1.6
rs13191362	6	*PARK2*	G/A	0.041 (0.112)	0.71677	0.00011	0.4
rs2033529	6	*TDRG1*	G/A	0.000 (0.082)	0.99922	0.00005	0.2
rs205262	6	*C6orf106*	G/A	0.063 (0.083)	0.44482	0.00001	0.0
rs2207139	6	*TFAP2B*	G/A	0.023 (0.099)	0.81892	0.00000	0.0
rs1167827	7	*HIP1*	G/A	0.098 (0.073)	0.18188	0.00026	1.0
rs2245368	7	*PMS2L11*	C/T	0.146 (0.101)	0.14578	0.00034	1.3
rs17405819	8	*HNF4G*	T/C	0.093 (0.081)	0.25063	0.00023	0.9
rs2033732	8	*RALYL*	C/T	0.177 (0.085)	0.03776	0.00088	3.4
rs10733682	9	*LMX1B*	A/G	0.037 (0.074)	0.61441	0.00021	0.8
rs10968576	9	*LINGO2*	G/A	0.067 (0.08)	0.40286	0.00015	0.6

rs1928295	9	*TLR4*	C/T	0.014 (0.073)	0.84496	0.00002	0.1
rs4740619	9	*C9orf93*	C/T	0.003 (0.074)	0.97016	0.00008	0.0
rs6477694	9	*EPB41L4B*	C/T	0.086 (0.076)	0.25859	0.00042	1.6
rs11191560	10	*NT5C2*	T/C	0.020 (0.134)	0.87886	0.00003	0.1
rs17094222	10	*HIF1AN*	C/T	0.115 (0.089)	0.20008	0.00027	1.0
rs7899106	10	*GRID1*	G/A	0.306 (0.171)	0.07277	0.00039	1.5
rs7903146	10	*TCF7L2*	C/T	0.014 (0.080)	0.85721	0.00010	0.4
rs12286929	11	*CADM1*	G/A	0.003 (0.074)	0.96457	0.00000	0.0
rs2176598	11	*HSD17B12*	T/C	0.114 (0.085)	0.17859	0.00060	2.3
rs3817334	11	*MTCH2*	T/C	0.041 (0.075)	0.58851	0.00007	0.3
rs4256980	11	*TRIM66*	G/C	0.069 (0.076)	0.36777	0.00023	0.9
rs11057405	12	*CLIP1*	G/A	0.161 (0.123)	0.19053	0.00017	0.7
rs7138803	12	*BCDIN3D*	A/G	0.091 (0.075)	0.22791	0.00033	1.3
rs12429545	13	*OLFM4*	G/A	0.125 (0.109)	0.25182	0.00054	2.1
rs10132280	14	*STXBP6*	C/A	0.008 (0.079)	0.92308	0.00002	0.1
rs11847697	14	*PRKD1*	T/C	0.126 (0.175)	0.47347	0.00001	0.0
rs12885454	14	*PRKD1*	C/A	0.014 (0.076)	0.85200	0.00000	0.0
rs7141420	14	*NRXN3*	T/C	0.050 (0.074)	0.49775	0.00021	0.8
rs3736485	15	*SCG3*	A/G	0.038 (0.074)	0.61107	0.00000	0.0
rs12446632	16	*GPRC5B*	G/A	0.147 (0.103)	0.15384	0.00060	2.3
rs1558902	16	*FTO*	A/T	0.250 (0.074)	0.00080	0.00185	7.2
rs758747	16	*NLRC3*	T/C	0.124 (0.082)	0.13120	0.00064	2.5
rs9925964	16	*KAT8*	A/G	0.042 (0.077)	0.58680	0.00015	0.6
rs1000940	17	*RABEP1*	A/G	0.064 (0.078)	0.41699	0.00023	0.9
rs12940622	17	*RPTOR*	G/A	0.157 (0.074)	0.03318	0.00141	5.5
rs6567160	18	*MC4R*	C/T	0.095 (0.087)	0.27557	0.00010	0.4
rs7243357	18	*GRP*	T/G	0.015 (0.097)	0.87527	0.00003	0.1
rs3810291	19	*ZC3H4*	A/G	0.137 (0.078)	0.07990	0.00048	1.9

Abbreviations: no., number; Chr., chromosome; se, standard error; F-stat, F-statistic.
[a]Allele increasing/allele decreasing: genetic associations reported with respect to the trait increasing allele.

Table I.6 Estimates (standard errors) and p-values of the genetic associations with the dual-energy X-ray (DXA) measurements for fat-free mass index for the 60 genetic variants selected under the liberal criteria based on 3,901 participants from UK Biobank. The R^2 value and F-statistic are provided from the model when the body composition measurement was regressed against the genetic variant.

rs no.	Chr.	Gene	Alleles[a]	Beta (se)	P-value	R^2	F-stat
rs11165643	1	*PTBP2*	T/C	0.002 (0.037)	0.94917	0.00000	0.0
rs12401738	1	*FUBP1*	A/G	0.025 (0.037)	0.50663	0.00002	0.1
rs12566985	1	*FPGT TNNI3K*	A/G	0.010 (0.037)	0.79035	0.00018	0.7
rs17024393	1	*GNAT2*	C/T	0.206 (0.121)	0.08918	0.00002	0.1
rs2820292	1	*NAV1*	C/A	0.019 (0.037)	0.59496	0.00007	0.3
rs543874	1	*SEC16B*	G/A	0.028 (0.045)	0.53926	0.00020	0.8
rs657452	1	*AGBL4*	G/A	0.016 (0.037)	0.66958	0.00017	0.6
rs1016287	2	*LINC01122*	C/T	0.015 (0.039)	0.70911	0.00001	0.1
rs11126666	2	*KCNK3*	A/G	0.034 (0.041)	0.41609	0.00005	0.2
rs11688816	2	*EHBP1*	A/G	0.012 (0.037)	0.73284	0.00003	0.1
rs1528435	2	*UBE2E3*	T/C	0.007 (0.037)	0.84171	0.00044	1.7
rs7599312	2	*ERBB4*	G/A	0.073 (0.041)	0.07376	0.00020	0.8
rs13078960	3	*CADM2*	G/T	0.068 (0.045)	0.13379	0.00002	0.1
rs1516725	3	*ETV5*	C/T	0.063 (0.053)	0.23308	0.00082	3.2
rs16851483	3	*RASA2*	T/G	0.056 (0.073)	0.44795	0.00006	0.2
rs2365389	3	*FHIT*	C/T	0.091 (0.037)	0.01361	0.00016	0.6
rs3849570	3	*GBE1*	A/C	0.005 (0.037)	0.89144	0.00010	0.4
rs6804842	3	*RARB*	G/A	0.031 (0.037)	0.39096	0.00018	0.7
rs10938397	4	*GNPDA2*	A/G	0.044 (0.037)	0.23556	0.00027	1.1
rs11727676	4	*HHIP*	T/C	0.008 (0.061)	0.89510	0.00015	0.6
rs17001654	4	*NUP54*	G/C	0.041 (0.051)	0.42731	0.00073	2.9
rs2112347	5	*POC5*	G/T	0.034 (0.037)	0.36312	0.00023	0.9
rs13191362	6	*PARK2*	A/G	0.053 (0.055)	0.34157	0.00081	3.2
rs2033529	6	*TDRG1*	A/G	0.013 (0.040)	0.74232	0.00029	1.1
rs205262	6	*C6orf106*	G/A	0.072 (0.041)	0.07896	0.00104	4.0
rs2207139	6	*TFAP2B*	G/A	0.045 (0.049)	0.35575	0.00063	2.5
rs1167827	7	*HIP1*	G/A	0.056 (0.036)	0.12278	0.00039	1.5
rs2245368	7	*PMS2L11*	C/T	0.056 (0.050)	0.26189	0.00016	0.6
rs17405819	8	*HNF4G*	T/C	0.050 (0.040)	0.21098	0.00021	0.8
rs2033732	8	*RALYL*	C/T	0.071 (0.042)	0.09143	0.00066	2.6
rs10733682	9	*LMX1B*	A/G	0.016 (0.036)	0.65733	0.00016	0.6
rs10968576	9	*LINGO2*	A/G	0.016 (0.039)	0.69000	0.00009	0.3

rs1928295	9	*TLR4*	T/C	0.022 (0.036)	0.55169	0.00021	0.8
rs4740619	9	*C9orf93*	T/C	0.048 (0.037)	0.18915	0.00026	1.0
rs6477694	9	*EPB41L4B*	C/T	0.077 (0.037)	0.03904	0.00010	0.4
rs11191560	10	*NT5C2*	C/T	0.029 (0.066)	0.65739	0.00009	0.3
rs17094222	10	*HIF1AN*	C/T	0.043 (0.044)	0.32529	0.00032	1.3
rs7899106	10	*GRID1*	G/A	0.024 (0.084)	0.77334	0.00000	0.0
rs7903146	10	*TCF7L2*	C/T	0.002 (0.039)	0.96940	0.00026	1.0
rs12286929	11	*CADM1*	A/G	0.009 (0.037)	0.81406	0.00000	0.0
rs2176598	11	*HSD17B12*	T/C	0.019 (0.042)	0.64603	0.00001	0.1
rs3817334	11	*MTCH2*	C/T	0.027 (0.037)	0.46186	0.00001	0.0
rs4256980	11	*TRIM66*	G/C	0.013 (0.038)	0.72791	0.00003	0.1
rs11057405	12	*CLIP1*	A/G	0.039 (0.061)	0.52044	0.00000	0.0
rs7138803	12	*BCDIN3D*	A/G	0.038 (0.037)	0.29981	0.00002	0.1
rs12429545	13	*OLFM4*	G/A	0.011 (0.054)	0.83867	0.00002	0.1
rs10132280	14	*STXBP6*	C/A	0.022 (0.039)	0.57731	0.00036	1.4
rs11847697	14	*PRKD1*	T/C	0.032 (0.086)	0.71241	0.00014	0.5
rs12885454	14	*PRKD1*	C/A	0.007 (0.038)	0.86129	0.00001	0.0
rs7141420	14	*NRXN3*	C/T	0.011 (0.036)	0.76915	0.00013	0.5
rs3736485	15	*SCG3*	G/A	0.001 (0.037)	0.97750	0.00016	0.6
rs12446632	16	*GPRC5B*	G/A	0.108 (0.051)	0.03419	0.00041	1.6
rs1558902	16	*FTO*	A/T	0.033 (0.037)	0.36684	0.00112	4.4
rs758747	16	*NLRC3*	T/C	0.045 (0.041)	0.26420	0.00009	0.4
rs9925964	16	*KAT8*	A/G	0.066 (0.038)	0.08290	0.00021	0.8
rs1000940	17	*RABEP1*	G/A	0.025 (0.039)	0.51302	0.00013	0.5
rs12940622	17	*RPTOR*	G/A	0.090 (0.036)	0.01351	0.00011	0.4
rs6567160	18	*MC4R*	C/T	0.108 (0.043)	0.01221	0.00116	4.5
rs7243357	18	*GRP*	T/G	0.005 (0.048)	0.92267	0.00004	0.2
rs3810291	19	*ZC3H4*	A/G	0.150 (0.038)	0.00009	0.00231	9.0

Abbreviations: no., number; Chr., chromosome; se, standard error; F-stat, F-statistic.
[a]Allele increasing/allele decreasing: genetic associations reported with respect to the trait increasing allele.

Table I.7 Estimates (standard errors) and p-values of the genetic associations with asthma (ever diagnosis and current asthma) for the 60 genetic variants selected under the liberal criteria based on 360,409 participants from UK Biobank.

				Ever		**Current**	
rs no.	Chr.	Gene	Allele[a]	Beta (se)	P-value	Beta (se)	P-value
rs11165643	1	*PTBP2*	T/C	0.016 (0.007)	0.0316	0.019 (0.009)	0.0404
rs12401738	1	*FUBP1*	A/G	0.008 (0.008)	0.3156	0.005 (0.009)	0.5985
rs12566985	1	*FPGT TNNI3K*	G/A	0.022 (0.007)	0.0029	0.018 (0.009)	0.0457
rs17024393	1	*GNAT2*	C/T	0.050 (0.023)	0.0308	0.050 (0.028)	0.0731
rs2820292	1	*NAV1*	C/A	0.001 (0.007)	0.8434	0.001 (0.009)	0.9297
rs543874	1	*SEC16B*	G/A	0.008 (0.009)	0.3737	0.006 (0.011)	0.5636
rs657452	1	*AGBL4*	G/A	0.005 (0.008)	0.5479	0.003 (0.009)	0.7191
rs1016287	2	*LINC01122*	T/C	0.007 (0.008)	0.4174	0.015 (0.010)	0.1349
rs11126666	2	*KCNK3*	G/A	0.013 (0.008)	0.1144	0.012 (0.010)	0.2559
rs11688816	2	*EHBP1*	G/A	0.004 (0.007)	0.5533	-0.001 (0.009)	0.9088
rs1528435	2	*UBE2E3*	T/C	0.011 (0.008)	0.1596	0.013 (0.009)	0.1442
rs7599312	2	*ERBB4*	A/G	0.018 (0.008)	0.0316	0.023 (0.010)	0.0264
rs13078960	3	*CADM2*	T/G	0.015 (0.009)	0.1077	0.021 (0.011)	0.0607
rs1516725	3	*ETV5*	C/T	0.001 (0.011)	0.8998	-0.012 (0.013)	0.3404
rs16851483	3	*RASA2*	G/T	0.016 (0.015)	0.2855	0.027 (0.018)	0.1390
rs2365389	3	*FHIT*	C/T	0.003 (0.008)	0.6543	0.008 (0.009)	0.4056
rs3849570	3	*GBE1*	C/A	0.001 (0.008)	0.8646	-0.007 (0.009)	0.4453
rs6804842	3	*RARB*	G/A	0.003 (0.007)	0.6894	0.006 (0.009)	0.5265
rs10938397	4	*GNPDA2*	G/A	0.017 (0.007)	0.0236	0.019 (0.009)	0.0313
rs11727676	4	*HHIP*	C/T	0.012 (0.012)	0.3320	0.026 (0.015)	0.0877
rs17001654	4	*NUP54*	G/C	0.016 (0.010)	0.1353	0.020 (0.013)	0.1214
rs2112347	5	*POC5*	T/G	0.005 (0.008)	0.5115	0.012 (0.009)	0.2042
rs13191362	6	*PARK2*	A/G	0.021 (0.011)	0.0585	0.021 (0.014)	0.1195
rs2033529	6	*TDRG1*	A/G	0.004 (0.008)	0.6072	0.004 (0.010)	0.6645
rs205262	6	*C6orf106*	A/G	0.012 (0.008)	0.1332	0.017 (0.010)	0.1026
rs2207139	6	*TFAP2B*	G/A	0.014 (0.010)	0.1520	0.018 (0.012)	0.1281
rs1167827	7	*HIP1*	G/A	0.011 (0.007)	0.1264	0.020 (0.009)	0.0294
rs2245368	7	*PMS2L11*	T/C	0.005 (0.010)	0.6085	0.000 (0.012)	0.9999
rs17405819	8	*HNF4G*	T/C	0.004 (0.008)	0.6523	-0.009 (0.010)	0.3695
rs2033732	8	*RALYL*	C/T	0.011 (0.008)	0.1944	0.013 (0.010)	0.2245
rs10733682	9	*LMX1B*	A/G	0.006 (0.007)	0.4500	0.005 (0.009)	0.6189
rs10968576	9	*LINGO2*	A/G	0.000 (0.008)	0.9579	0.003 (0.010)	0.7835
rs1928295	9	*TLR4*	T/C	0.012 (0.007)	0.1083	0.013 (0.009)	0.1566
rs4740619	9	*C9orf93*	T/C	0.015 (0.007)	0.0365	0.014 (0.009)	0.1096
rs6477694	9	*EPB41L4B*	C/T	0.011 (0.008)	0.1560	0.010 (0.009)	0.3075
rs11191560	10	*NT5C2*	C/T	0.018 (0.014)	0.1951	0.030 (0.017)	0.0710
rs17094222	10	*HIF1AN*	C/T	0.004 (0.009)	0.6817	0.009 (0.011)	0.4307

rs7899106	10	*GRID1*	G/A	0.035 (0.017)	0.0357	0.036 (0.020)	0.0764
rs7903146	10	*TCF7L2*	C/T	0.005 (0.008)	0.5245	0.003 (0.010)	0.7342
rs12286929	11	*CADM1*	G/A	0.001 (0.007)	0.8617	0.011 (0.009)	0.2265
rs2176598	11	*HSD17B12*	T/C	0.001 (0.009)	0.9038	0.005 (0.010)	0.6121
rs3817334	11	*MTCH2*	T/C	0.000 (0.007)	0.9514	0.003 (0.009)	0.7513
rs4256980	11	*TRIM66*	C/G	0.000 (0.008)	0.9985	-0.001 (0.009)	0.9328
rs11057405	12	*CLIP1*	G/A	0.020 (0.012)	0.0904	0.010 (0.015)	0.4750
rs7138803	12	*BCDIN3D*	G/A	0.002 (0.008)	0.7583	0.003 (0.009)	0.7842
rs12429545	13	*OLFM4*	A/G	0.003 (0.011)	0.7918	0.000 (0.013)	0.9705
rs10132280	14	*STXBP6*	C/A	0.009 (0.008)	0.2422	0.007 (0.010)	0.4546
rs11847697	14	*PRKD1*	T/C	0.017 (0.018)	0.3347	0.008 (0.022)	0.7218
rs12885454	14	*PRKD1*	A/C	0.003 (0.008)	0.6737	0.007 (0.009)	0.4310
rs7141420	14	*NRXN3*	T/C	0.015 (0.007)	0.0478	0.014 (0.009)	0.1271
rs3736485	15	*SCG3*	A/G	0.015 (0.007)	0.0396	0.020 (0.009)	0.0231
rs12446632	16	*GPRC5B*	G/A	0.000 (0.011)	0.9775	0.012 (0.013)	0.3642
rs1558902	16	*FTO*	A/T	0.016 (0.007)	0.0340	0.028 (0.009)	0.0024
rs758747	16	*NLRC3*	C/T	0.018 (0.008)	0.0312	0.013 (0.010)	0.1979
rs9925964	16	*KAT8*	A/G	0.006 (0.008)	0.4129	0.006 (0.009)	0.4909
rs1000940	17	*RABEP1*	G/A	0.003 (0.008)	0.6763	0.001 (0.010)	0.9357
rs12940622	17	*RPTOR*	G/A	0.005 (0.007)	0.4673	0.008 (0.009)	0.3582
rs6567160	18	*MC4R*	C/T	0.005 (0.009)	0.5723	0.005 (0.011)	0.6676
rs7243357	18	*GRP*	G/T	0.021 (0.010)	0.0269	0.019 (0.012)	0.1120
rs3810291	19	*ZC3H4*	A/G	0.023 (0.008)	0.0035	0.030 (0.010)	0.0016

Abbreviations: no., number; Chr., chromosome; se, standard error.
[a]Allele increasing/allele decreasing: genetic associations reported with respect to the trait increasing allele under the primary definition.

I.2 Two-sample Mendelian randomization study

I.2.1 Genetic associations with asthma from the GABRIEL consortium

Table I.8 contains the genetic associations with asthma for the 42 liberal variants extracted from the GABRIEL consortium. The rs number and r^2 value (measure of linkage disequilibrium) of the proxy variant are given when relevant. The final column in Table I.8 indicates whether the variant was included in the conservative set.

Table I.8 Estimates (standard errors) and p-values of the genetic associations with asthma extracted from the GABRIEL consortium. The rs number and r^2 value (measure of linkage disequilibrium) of the proxy variant are given when relevant. The final column indicates whether the variant was included in the conservative set.

rs no.	Proxy no.	Chr.	Gene	r^2	Minor/ major	MAF	Effect allele	Estimate (se)	P-value	Con
rs11165643		1	*PTBP2*		C/T	0.491	C	-0.003 (0.020)	0.880	Yes
rs12401738	rs4130548	1	*FUBP1*	1.000	C/T	0.422	C	-0.014 (0.021)	0.502	No
rs12566985	rs6604872	1	*FPGT-TNNI3K*	1.000	T/C	0.394	C	0.001 (0.021)	0.951	Yes
rs2820292	rs1032524	1	*NAV1*	0.894	T/C	0.399	T	0.018 (0.020)	0.357	No
rs657452		1	*AGBL4*		G/A	0.432	A	0.007 (0.021)	0.751	Yes
rs11126666		2	*KCNK3*		A/G	0.447	A	-0.014 (0.022)	0.530	No
rs11688816		2	*EHBP1*		G/A	0.417	A	0.012 (0.020)	0.545	Yes
rs7599312		2	*ERBB4*		A/G	0.327	A	0.030 (0.023)	0.191	Yes
rs1516725	rs6809651	3	*ETV5*	1.000	A/G	0.172	G	0.021 (0.030)	0.473	No
rs16851483	rs3821709	3	*RASA2*	1.000	C/T	0.117	T	0.002 (0.040)	0.965	No
rs3849570	rs7620240	3	*GBE1*	0.991	T/G	0.270	G	-0.037 (0.021)	0.074	Yes
rs6804842		3	*RARB*		G/A	0.487	A	-0.007 (0.020)	0.717	Yes
rs10938397	rs12641981	4	*GNPDA2*	0.992	T/C	0.359	C	-0.010 (0.020)	0.608	No
rs17001654	rs17001561	4	*NUP54*	0.952	A/G	0.229	G	-0.070 (0.027)	0.010	Yes
rs2033529		6	*TDRG1*		G/A	0.241	G	0.002 (0.022)	0.941	Yes
rs205262		6	*C6orf106*		G/A	0.224	G	-0.005 (0.022)	0.811	No
rs2207139	rs943005	6	*TFAP2B*	1.000	T/C	0.101	C	-0.011 (0.026)	0.683	Yes
rs1167827		7	*HIP1*		G/A	0.458	A	-0.043 (0.020)	0.033	No
rs10733682		9	*LMX1B*		G/A	0.403	A	0.007 (0.020)	0.708	Yes
rs10968576		9	*LINGO2*		G/A	0.263	A	-0.015 (0.021)	0.489	Yes
rs1928295		9	*TLR4*		T/C	0.417	C	-0.003 (0.020)	0.894	No
rs4740619		9	*C9orf93*		T/C	0.488	C	0.030 (0.020)	0.134	No
rs6477694		9	*EPB41L4B*		T/C	0.486	C	0.002 (0.021)	0.937	Yes
rs17094222	rs17113301	10	*HIF1AN*	0.903	A/C	0.277	A	0.065 (0.024)	0.007	Yes
rs7899106	rs11201714	10	*GRID1*	1.000	A/C	0.073	C	-0.012 (0.047)	0.793	Yes
rs7903146		10	*TCF7L2*		T/C	0.137	C	-0.010 (0.022)	0.663	Yes
rs12286929	rs12421648	11	*CADM1*	0.821	G/A	0.429	G	0.033 (0.020)	0.099	No
rs3817334	rs7124681	11	*MTCH2*	0.992	A/C	0.456	C	-0.023 (0.020)	0.266	Yes
rs4256980	rs2316901	11	*TRIM66*	0.996	A/G	0.407	A	0.015 (0.021)	0.479	Yes
rs11057405		12	*CLIP1*		A/G	0.154	A	-0.005 (0.034)	0.883	Yes
rs7138803		12	*BCDIN3D*		A/G	0.469	A	0.047 (0.020)	0.021	Yes
rs12429545		13	*OLFM4*		A/G	0.162	A	-0.046 (0.030)	0.129	Yes
rs10132280	rs8015400	14	*STXBP6*	0.890	C/A	0.289	C	0.003 (0.021)	0.904	Yes
rs12885454	rs1307813	14	*PRKD1*	0.996	C/A	0.285	C	-0.004 (0.021)	0.851	Yes
rs7141420		14	*NRXN3*		T/C	0.467	C	-0.026 (0.020)	0.196	No
rs12446632		16	*GPRC5B*		A/G	0.192	A	-0.036 (0.029)	0.211	Yes
rs758747		16	*NLRC3*		T/C	0.220	C	0.000 (0.022)	0.989	No
rs9925964	rs889548	16	*KAT8*	1.000	T/C	0.292	C	-0.022 (0.021)	0.294	Yes
rs1000940		17	*RABEP1*		G/A	0.189	G	-0.012 (0.022)	0.566	Yes
rs12940622		17	*RPTOR*		A/G	0.500	A	0.003 (0.020)	0.885	Yes
rs6567160	rs571312	18	*MC4R*	1.000	A/C	0.297	A	-0.007 (0.023)	0.759	Yes
rs7243357	rs9961404	18	*GRP*	0.872	T/C	0.113	T	-0.036 (0.027)	0.190	Yes

Abbreviations: no., number; Chr., chromosome; MAF, minor allele frequency; se, standard error; Con., conservative.

I.2.2 Genetic associations with BMI from UK Biobank

Table I.9 contains the estimates of the genetic associations with the 19 proxy variants used from the GABRIEL consortium for BMI, FMI (BIA and DXA) and FFMI (BIA and DXA) based on the 360,409 participants from UK Biobank. The genetic associations were adjusted for the first 10 PCs, gender, and height.

Table I.9 Estimates (standard errors) and p-values of the genetic associations with body mass index, fat mass index (BIA and DXA) and fat-free mass index (BIA and DXA) for the 19 proxy genetic variants used in the two-sample Mendelian randomization analysis based on the 360,409 participants from UK Biobank.

				BMI		**FMI** (BIA)		**FFMI** (BIA)		**FMI** (DXA)		**FFMI** (DXA)	
rs no.	Proxy no.	Chr.	Alleles[a]	Estimate (se)	P-value	Estimate (se)	P-value	Estimate (se)	P-value	Estimate (se)	P-value	Estimate (se)	P-value
rs2820292	rs1032524	1	C/T	0.090 (0.011)	4.26×10^{-16}	0.056 (0.008)	3.50×10^{-13}	0.034 (0.004)	1.26×10^{-16}	0.052 (0.073)	0.480	0.019 (0.036)	0.600
rs12401738	rs4130548	1	C/T	0.090 (0.011)	2.48×10^{-15}	0.054 (0.008)	1.33×10^{-11}	0.036 (0.004)	5.60×10^{-17}	0.111 (0.075)	0.137	0.027 (0.037)	0.469
rs12566985	rs6604872	1	C/T	-0.090 (0.011)	6.78×10^{-16}	-0.053 (0.008)	7.33×10^{-12}	-0.037 (0.004)	5.81×10^{-18}	0.028 (0.075)	0.712	0.010 (0.037)	0.791
rs16851483	rs3821709	3	C/T	0.162 (0.022)	2.88×10^{-13}	0.109 (0.016)	2.41×10^{-12}	0.054 (0.008)	2.67×10^{-10}	-0.096 (0.148)	0.517	0.054 (0.073)	0.458
rs1516725	rs6809651	3	A/G	-0.148 (0.016)	2.76×10^{-20}	-0.082 (0.011)	2.97×10^{-13}	-0.067 (0.006)	1.89×10^{-27}	0.044 (0.107)	0.679	-0.065 (0.053)	0.222
rs3849570	rs7620240	3	T/G	0.047 (0.012)	4.65×10^{-5}	0.036 (0.008)	9.86×10^{-6}	0.011 (0.004)	0.010	-0.061 (0.076)	0.417	0.004 (0.037)	0.916
rs10938397	rs12641981	4	T/C	0.147 (0.011)	1.42×10^{-39}	0.100 (0.008)	1.41×10^{-37}	0.047 (0.004)	1.83×10^{-28}	-0.006 (0.075)	0.935	-0.044 (0.037)	0.234
rs17001654	rs17001561	4	A/G	0.071 (0.015)	4.19×10^{-6}	0.041 (0.011)	0.001	0.030 (0.006)	3.18×10^{-7}	0.103 (0.102)	0.310	0.018 (0.050)	0.716
rs2207139	rs943005	6	T/C	0.187 (0.015)	5.19×10^{-37}	0.117 (0.010)	7.02×10^{-30}	0.07 (0.006)	5.06×10^{-36}	0.019 (0.098)	0.847	0.044 (0.049)	0.369
rs7899106	rs11201714	10	A/C	0.141 (0.025)	2.71×10^{-8}	0.080 (0.018)	7.23×10^{-6}	0.062 (0.01)	2.02×10^{-10}	0.306 (0.171)	0.073	0.024 (0.084)	0.778
rs17094222	rs17113301	10	A/C	0.067 (0.013)	4.64×10^{-7}	0.049 (0.009)	1.91×10^{-7}	0.019 (0.005)	0.002	0.118 (0.088)	0.183	0.026 (0.044)	0.547
rs12286929	rs12421648	11	A/G	0.062 (0.011)	3.15×10^{-8}	0.037 (0.008)	1.85×10^{-6}	0.025 (0.004)	7.36×10^{-9}	0.024 (0.074)	0.749	-0.008 (0.037)	0.832
rs4256980	rs2316901	11	A/G	-0.074 (0.012)	1.79×10^{-10}	-0.053 (0.008)	9.23×10^{-11}	-0.022 (0.004)	1.16×10^{-6}	-0.076 (0.076)	0.322	-0.010 (0.038)	0.788
rs3817334	rs7124681	11	A/C	0.108 (0.011)	5.88×10^{-22}	0.087 (0.008)	2.01×10^{-28}	0.021 (0.004)	5.33×10^{-7}	0.037 (0.075)	0.619	-0.029 (0.037)	0.427
rs12885454	rs1307813	14	A/C	0.079 (0.012)	7.74×10^{-12}	0.050 (0.008)	4.43×10^{-10}	0.029 (0.004)	7.18×10^{-11}	0.014 (0.076)	0.854	0.004 (0.038)	0.918
rs10132280	rs8015400	14	A/C	0.105 (0.012)	6.01×10^{-19}	0.057 (0.008)	3.53×10^{-12}	0.048 (0.005)	3.98×10^{-26}	0.055 (0.078)	0.478	0.038 (0.038)	0.322
rs9925964	rs889548	16	T/C	-0.127 (0.012)	2.78×10^{-28}	-0.089 (0.008)	1.50×10^{-28}	-0.038 (0.004)	6.40×10^{-18}	-0.042 (0.077)	0.590	-0.067 (0.038)	0.081
rs6567160	rs571312	18	A/C	0.277 (0.013)	2.10×10^{-100}	0.160 (0.009)	4.95E-69	0.117 (0.005)	2.70×10^{-123}	0.082 (0.087)	0.346	0.108 (0.043)	0.012
rs7243357	rs9961404	18	T/C	-0.099 (0.015)	2.51×10^{-11}	-0.059 (0.010)	1.30×10^{-8}	-0.04 (0.006)	1.29×10^{-12}	0.072 (0.099)	0.468	-0.005 (0.049)	0.919

Abbreviations: BMI, body mass index; FMI, fat mass index; FFMI, fat-free mass index; BIA, bioelectrical impedance analysis; DXA, dual X-ray emission absorptiometry; no., number; Chr., chromosome; se, standard error.
[a]Effect allele/non-effect allele.